
This is the **published version** of the bachelor thesis:

Cuesta Guiu, Adrià; Martínez Quirante, Roser, dir. El biaix de la IA i els LAWS des d'una perspectiva jurídica. 2021. 75 pag. (994 Grau en Administració i Direcció d'Empreses i Grau en Dret)

This version is available at <https://ddd.uab.cat/record/245315>

under the terms of the  license



**Universitat Autònoma
de Barcelona**

FACULTAT DE DRET

EL BIAIX DE LA IA I ELS LAWS DES D'UNA PERSPECTIVA JURÍDICA

TREBALL DE FI DE GRAU

AUTOR: ADRIÀ CUESTA GUIU

**DOBLE GRAU EN ADMINISTRACIÓ I DIRECCIÓ
D'EMPRESES I DRET**

CURS 2020 – 2021

DIRECTORA: ROSER MARTÍNEZ QUIRANTE

DATA D'ENTREGA: 11 DE MAIG DE 2021

«La preocupació per la militarització de la intel·ligència artificial va en augment. La possibilitat que hi hagi armes que puguin seleccionar i atacar un objectiu per si soles dispara moltes alarmes i podria provocar noves curses armamentistes. (...) Expliquem-ho tal com és. La possibilitat que hi hagi màquines amb discreció i poder per endur-se vides humanes és moralment repugnant.»

António Guterres, Secretari General de les Nacions Unides.

AGRAÏMENTS

En primer lloc, agrair a la meva tutora Roser Martínez Quirante que durant la seva docència de les assignatures de Dret Administratiu ens introduís en el món dels LAWS i els moviments internacionals que s'estaven desplegant. Fent-me interessar, encara més, per aquesta situació amb la presentació i posterior lectura del seu llibre Inteligencia Artificial y armas letales autónomas, un nuevo reto para Naciones Unidas. També per tota la ajuda i *expertise* proporcionada tant de forma directa com a través dels seus llibres.

I en segon lloc, a tota la meva família i amics que han contribuït al no desesperar i tirar endavant amb el treball. I gràcies també a la seva paciència i suport incondicional, fent costat durant tots aquests mesos.

RESUM

La IA ens posiciona front a un dels majors avenços de la història de la humanitat en poder comptar amb tecnologies tant o més intel·ligents que els éssers humans. Tot i esdevenir un escenari d'avantatges increïbles, compta amb una sèrie d'inconvenients que poden vulnerar greument certs principis o drets humans internacionals, un dels més destacats és el biaix algorítmic de la IA. Un dels majors riscos d'aquesta intel·ligència és la seva militarització incorporant-la en els LAWS o *killer robots*, deixant que aquests robots abatin objectius al seu lliure arbitri. Infringint així els principis més fonamentals del Dret Humanitari Internacional.

La incorporació immediata de la IA en el nostre dia a dia sense comptar amb una regulació estricta planteja tota una sèrie de qüestions jurídiques que fins ara la societat i els actors legislatius no havien previst. Aquesta falta de regulació és més perillosa encara en el disseny, la construcció i l'ús dels LAWS per les seves accions letals. És per això que a nivell internacional s'està intentant arribar a un marc jurídic que reguli aquestes noves tecnologies, i pel que en aquest treball s'ha intentat abordar els aspectes claus a tenir en compte en les futures regulacions respecte al biaix de la IA i la limitació en l'ús dels LAWS.

Paraules clau: IA, Biaix algorítmic, LAWS, Supervisió humana significativa, Dret Internacional Humanitari

ÍNDEX

1. INTRODUCCIÓ	1
2. INTEL·LIGÈNCIA ARTIFICIAL	2
2.1. MODELS PREDICTIUS.....	4
2.2. BIAIX ALGORÍTMIC	6
3. ASPECTES FONAMENTALS A REGULAR RESPECTE LA IA	9
3.1. REGULACIONS EN EL DISSENY I DESENVOLUPAMENT DE IA.....	9
3.1.1. INCORPORACIÓ DE VALORS I PRINCIPIS ÈTICS EN L'ALGORITME	11
3.1.2. CODIS DEONTOLÒGICS PROFESSIONALS EN EL DISSENY I DESENVOLUPAMENT DE IA	14
3.2. REGULACIÓ DE L'ÚS EFECTIU DE LA TECNOLOGIA IA	15
3.2.1. AUDITORIES I CONTROL EXTERN	16
3.2.2. RESPONSABILITAT CIVIL.....	17
3.2.3. ASSEGURANÇA DE LA RESPONSABILITAT CIVIL.....	20
4. ROBOTS AUTÒNOMS.....	21
5. LETHAL AUTONOMOUS WEAPONS SYSTEMS (LAWS).....	24
5.1. RISCOS DELS LAWS.....	26
5.1.1. PERILLS NO LEGALS.....	28
5.1.2. PERILLS LEGALS	30
6. ASPECTES FONAMENTALS A REGULAR RESPECTE ELS LAWS	33
6.1. REGULACIÓ PROHIBITIVA.....	33
6.2. REGULACIÓ LIMITADORA.....	38
6.2.1. LLEIS DE LA ROBÒTICA D'ASIMOV	39
6.2.2. SUPERVISIÓ HUMANA	39
6.2.3. NO LETALITAT I KILL SWITCH BUTTON	44
6.2.4. RÈGIM DE RESPONSABILITAT	47
7. CONCLUSIÓ.....	55
BIBLIOGRAFIA	62

1. INTRODUCCIÓ

En el treball es busca plantejar com la comunitat internacional intenta fer front a l'actual risc que suposa la implementació i el desenvolupament de noves tecnologies amb intel·ligència artificial (IA) integrada. Es tractarà de donar solució a problemes derivats del biaix algorítmic amb la regulació en vigor actual, tant nacional com internacional. Però en vista de l'actual falta de normativa, s'intentaran establir i desenvolupar els principals aspectes o punts a tractar per les futures regulacions.

Una de les majors implicacions en aquest estudi és l'actual desregulació, tant en el sector de la IA com pel que fa als Lethal Autonomous Weapon Systems (LAWS). Aquesta falta de normativa a nivell internacional, i donades les grans vulneracions a drets humans universals o fonamentals europeus, comporta una certa urgència en l'establiment d'un marc normatiu regulador. És per això que en aquest treball s'hi podran trobar propostes en aquest sentit, seguint sempre moviments doctrinals o institucionals, i de vegades socials, per entendre quins serien els punts de vital importància a regular. En el cas dels LAWS s'estudiaran més en concret les implicacions que té la seva utilització respecte diverses branques del Dret, però sobretot les vulneracions al Dret Internacional Humanitari.

Pel que fa a l'estructura, el treball està comprès per dos grans blocs. Aquests blocs es corresponen a les fases lògiques de desenvolupament per poder entendre al final de cada un d'ells la matèria tractada, podent així comprendre el funcionament del següent bloc. És a dir, en el primer bloc es tracta la intel·ligència artificial, que és del tot necessari conèixer el seu funcionament per poder assimilar el funcionament dels LAWS, que constitueixen el segon bloc del treball. En cada un dels blocs no només es defineixen els conceptes claus, sinó que es tracten en profunditat els riscos associats amb cada un d'ells i els aspectes principals amb els que iniciar les futures propostes regulatòries.

2. INTEL·LIGÈNCIA ARTIFICIAL

Encara que la intel·ligència artificial (IA) no té una definició estàndard, una de les més utilitzades la descriu com un camp de la ciència i la enginyeria que s'encarrega de la comprensió, des d'un punt de vista informàtic, del que es denomina comportament intel·ligent, ocupant-se, també, de la creació d'artefactes que exhibeixen aquest comportament.¹ Es tracta d'un concepte ampli que tracta les aspiracions dels sectors tecnocientífics i l'adveniment d'un sistema econòmic, sociològic, antropològic, filosòfic i legal totalment nou.² Doncs, és l'estudi de qualsevol dispositiu que percebi l'entorn i prengui accions per tal d'assolir uns objectius en adaptar-se a situacions no prèviament conegudes i aprenent de l'experiència, tenint en compte que la tecnologia no és intel·ligent *per se*, sinó que simula tenir-ne.³ La intel·ligència artificial està integrada per diversos processos que permeten que pugui percebre l'entorn i interactuar amb el mateix, són l'aprenentatge, el raonament, la resolució de problemes, la percepció i l'ús del llenguatge.

La IA requereix inicialment d'unes instruccions, prèviament programades, que concretin les accions ha realitzar per resoldre un determinat problema, el que es coneix com estructura algorítmica.⁴ Aquest algoritme pot arribar a aprendre de les dades i pot predir comportaments futurs a través d'aquestes, mitjançant l'aplicació del Machine Learning. El que vol dir que és el propi sistema el que es millora a si

¹ PINO DíEZ, Raúl; GÓMEZ GÓMEZ, Alberto; ABAJO MARTÍNEZ, Nicolás de. 2001. *Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva*. Universidad de Oviedo : Servicio de publicaciones, 2001.

² MARTÍNEZ QUIRANTE, Roser; RÓDRIGUEZ ÁLVAREZ, Joaquín. 2018. *Inteligencia artificial y armas letales autónomas*. Gijón : Ediciones TREA, 2018.

³ MARTÍNEZ QUIRANTE, Roser; RÓDRIGUEZ ÁLVAREZ, Joaquín. 2019. *Towards a new AI race. The challenge of lethal autonomous weapons systems (LAWS) for the United Nations*. Pamplona : Editorial Aranzadi, 2019.

⁴ NAVAS NAVARRO, Susana, et al. 2017. *Inteligencia artificial. Tecnología. Derecho*. Valencia : Tirant lo Blanch, 2017.

mateix, de forma autònoma sense intervenció humana.⁵ El major rendiment i eficiència en el processament de grans quantitats de dades s'obté en la unió del Machine Learning amb les tecnologies cognitives i la IA. En aquest context sorgeixen els robots cognitius que són aquells que compleixen un objectiu mitjançant la percepció de l'entorn, prenent atenció als esdeveniments importants, planejant el què fer, anticipant els resultats de les seves accions o d'altres agents i aprenent de la interacció resultant. Doncs, fan front a la incertesa d'entorns naturals amb un continu aprenentatge, raonament i compartint el seus coneixements.⁶

La IA es cataloga en funció de la seva capacitat per imitar característiques humanes, la tecnologia emprada i les seves aplicacions. Seguint aquesta estructura es diferencien tres estadis de IA.⁷

1. La *narrow AI* o feble en tenir molt limitades les seves capacitats.
2. La *Strong* o *Deep AI* en tenir capacitat per imitar la intel·ligència humana i el seu comportament.
3. La *Artificial Superintelligence* és el concepte hipotètic en el que la tecnologia de la IA supera les habilitats i la pròpia intel·ligència humana.

Actualment, degut a les limitacions tecnològiques, només hem assolit el primer estadi o *narrow AI*, ja que és aquella que està orientada a la realització de tasques determinades, per a les que ha estat programat el seu algoritme. Encara que sent tecnologia en IA de nivell més inferior, ja és capaç d'aprendre per si mateixa mitjançant el Machine Learning o el Deep Learning, però de forma limitada a través de l'anàlisi de gran quantitat de dades. Serien exemples els assistents virtuals com

⁵ GONZÁLEZ, Andrés. 2019. ¿Qué es Machine Learning? *Cleverdata*. [En línia] 2019. [Data: 3 / Gener / 2021.] <https://cleverdata.io/que-es-machine-learning-big-data/>.

⁶ TECHNICAL COMMITTEE FOR COGNITIVE ROBOTS. IEEE Robotics & Automation Society. [En línia] [Data: 3 / Gener / 2021.] <https://www.ieee-ras.org/cognitive-robotics>.

⁷ O'CARROLL, Brodie. 2017. What are the three types of AI? A guide to narrow, general, and super artificial intelligence. *Codebots*. [En línia] 24 / Octubre / 2017. [Data: 3 / Febrer / 2021.] <https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible>.

ara Siri, Alexa o Cortana, o les tecnologies de reconeixement facial, de veu o imatge.

Els estadis *Strong AI* i *Artificial Superintelligence* no estan desenvolupats en l'actualitat, encara que s'esperen que en un futur proper es realitzin nous avenços considerables en la matèria. En la primera, la IA profunda, és aquella que imita a la perfecció la intel·ligència humana i el seu comportament, de manera que dificulta la diferenciació entre humà i tecnologia segons el *Test de Turing*.⁸ El concepte de superintel·ligència artificial esdevé un escenari hipotètic en el qual la tecnologia de IA no és només capaç d'imitar les experiències i emocions humanes, sinó que expressa emocions, necessitats, creences i desitjos per si mateixa.⁹

No s'ha de pensar, observant les classificacions anteriors, que en trobar-nos en un estadi infantil de la IA, en ser el més limitat, es troba aquest lliure de riscos. Doncs, en la programació de les ordres més simples per poder dur a terme aquelles determinades accions es poden veure compromesos certs principis o drets, per la possible falta de neutralitat del propi algoritme com de la tecnologia. Originant-se aquesta afectació en la pròpia programació inicial de l'algoritme o en la posterior mineria de dades que realitzi la pròpia IA.

2.1. MODELS PREDICTIUS

Una de les funcions primordials que en l'actualitat realitza la tecnologia IA és la creació o elaboració de perfils o models predictius. L'elaboració d'aquests perfils és qualsevol forma de tractament automatitzat de dades personals per avaluar determinats aspectes personals d'una persona física, en particular, per analitzar o predir aspectes relatius al rendiment professional, situació econòmica, salut,

⁸ MARTÍNEZ QUIRANTE, et al., 2018, *op. cit.*

⁹ O'CARROLL, 2017, *op. cit.*

preferències personals, interessos, fiabilitat, comportament, ubicació o moviment de la persona objecte del perfil.¹⁰

La incorporació, doncs, del Machine Learning a les tecnologies de IA permeten que mitjançant el Big Data es generin unes bases de dades que es poden explotar amb la fi de preveure futurs comportament.¹¹ La incorporació del Big Data junt amb la IA en àmbits tant comuns o habituals com són les xarxes socials ha permès analitzar els grans volums de dades que es generen en l'ús d'aquestes plataformes, creant models predictius molt més encertats dels desenvolupats pel màrqueting tradicional.¹² Aquests models són utilitzats per predir amb gran precisió els productes o tipus d'anunci a mostrar a cada un dels usuaris, la franja horària en la que mostrar-ho per augmentar la probabilitat de compra o *engagement* o, fins i tot, els continguts a crear per part de companyies de *streaming*. Un exemple de com la IA crea models predictius és l'estudi realitzat per IBM en el que es demostra com l'anàlisi de les tuitades i missatges de Facebook poden definir la personalitat de la persona, podent conèixer si deixarà de ser soltera i es casarà, o fins i tot característiques de la personalitat com l'altruisme, el pensament radical, la depressió, etc.¹³

El problema que presenten aquests models predictius individualitzats és que les decisions són adoptades per agents autònoms que les prenen de manera automatitzada basada en els perfils creats pels propis algoritmes, afectant, en la

¹⁰ PARLAMENT EUROPEU I CONSELL (art.4.4). 2016. REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga. Bruselas : DOUE-L-2016-80807., 2016.

¹¹ GONZÁLEZ, 2019, *op. cit.*

¹² ACTIONSDATA. 2020. Big Data: Cómo lo usan las redes sociales. *ActionsData*. [En línia] 30 / Gener / 2020. [Data: 3 / Febrer / 2021.] <https://www.actionsdata.com/blog/big-data-como-lo-usan-las-redes-sociales>.

¹³ BBVA COMMUNICATIONS. 2017. El poder predictivo de las redes sociales. *BBVA*. [En línia] 9 / Agost / 2017. [Data: 3 / Febrer / 2021.] <https://www.bbva.com/es/poder-predictivo-redes-sociales/>.

majoria dels casos, de manera important, en el tractament de les dades del propi interessat, sense ser conscient habitualment d'aquesta afectació perjudicial i discriminatòria.¹⁴

2.2. BIAIX ALGORÍTMIC

La IA, com tota nova tecnologia, presenta tota una sèrie d'avantatges en tots els nivells, tant productius com socials, però també comporta tota una sèrie de riscos.¹⁵ La Unió Europea (UE) ja n'és conscient sobre els riscos que planteja la implementació de la IA en l'actualitat, és per això que al febrer de l'any 2020 va adoptar el "LIBRO BLANCO sobre inteligencia artificial – un enfoque europeo orientado a la excelencia i la confianza". En aquest Llibro Blanco es tracten, entre d'altres necessitats, la obligació per part dels desenvolupadors i implementadors de la IA d'estar subjectes a la legislació europea en matèria de drets fonamentals, protecció dels consumidors i normes sobre responsabilitat dels productes i responsabilitat civil. Algunes característiques de la IA, com poden ser l'opacitat o falta de neutralitat, poden comportar efectes imprevistos o que es puguin emprar per a fins malintencionats.¹⁶ Però, no és únicament la Unió qui preveu o té en compte un possible biaix o opacitat en la programació de la tecnologia de IA, sinó que diversos autors ja venen advertint. És exemple la pròpia descripció que fa en Kentaro Toyama sobre la tecnologia, sent aquesta un amplificador de la voluntat humana, doncs, capaç d'actuar seguint el biaix dels seus creadors o usuaris.¹⁷

Pel que fa al biaix algorítmic s'haurien de tractar els orígens que condueixen a l'existència del biaix, doncs, no és únicament conseqüència de prejudicis en les dades d'entrenament. El biaix pot provenir del moment inicial, abans de tenir dades,

¹⁴ NAVAS NAVARRO, et al., 2017, *op. cit.*

¹⁵ COMISSIÓ EUROPEA, 2020. *Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*. Bruselas : s.n., 2020. COM(2020) 65 final.

¹⁶ *Ibid.*

¹⁷ MARTÍNEZ QUIRANTE, et al., 2019, *op. cit.*

en el que s'estableix l'objectiu del model d'aprenentatge; de la mineria de dades amb dades no representatives de la realitat o amb prejudicis existents; o amb la introducció de determinats biaixos en el moment de preparar les dades i seleccionar els atributs que l'algoritme tingui en compte.¹⁸ Són diverses les investigacions recents les que han detectat que els grups històricament desfavorits, com són les persones negres i les dones, pateixen de forma més notòria els efectes del biaix algorítmic, que en contra, els grups dominants es beneficien de majors índexs de precisió en front al grups minoritaris.¹⁹

Les conseqüències que comporten la utilització d'aquestes tecnologies esbiaixades ja ha tingut ressò en els últims anys en grans companyies com ara Facebook o Amazon, per esmentar dues grans empreses. En el cas del gegant tecnològic de les xarxes socials, Facebook, han sigut varis els escàndols relacionats amb les tecnologies IA que desenvolupen, doncs, des de sistemes intel·ligents negociadors que van desenvolupar el seu propi idioma derivat de l'anglès que només comprenia el propi sistema²⁰, o discriminant en funció del gènere, raça o religió en la decisió autònoma de l'algoritme per mostrar uns o altre anuncis.²¹ Companyies com Amazon van desenvolupar i incorporar la tecnologia IA en el sistema de contractació per tal de puntuar les candidatures de treball en funció dels seus

¹⁸ HAO, Karen. 2019. Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo. *MIT Technology Review*. [En línia] Technology Review Inc., 8 / Febrer / 2019. [Data: 3 / Febrer / 2021.] <https://www.technologyreview.es/s/10924/como-se-produce-el-sesgo-algoritmico-y-por-que-es-tan-dificil-detenerlo>.

¹⁹ MARTÍNEZ, Naroa; MATUTE, Helena. 2020. DISCRIMINACIÓN RACIAL EN LA INTELIGENCIA ARTIFICIAL. *Ethic*. [En línia] 10 / Setembre / 2020. [Data: 3 / Febrer / 2021.] <https://ethic.es/2020/09/discriminacion-racial-en-la-inteligencia-artificial-sesgos/>.

²⁰ TECH2 NEWS STAFF. 2017. FACEBOOK RESEARCHERS SHUT DOWN AI BOTS THAT STARTED SPEAKING IN A LANGUAGE UNINTELLIGIBLE TO HUMANS. *TECH2*. [En línia] Firstpost, 1 / Agost / 2017. [Data: 4 / Febrer / 2021.] <https://www.firstpost.com/tech/news-analysis/facebook-researchers-shut-down-ai-bots-that-started-speaking-in-a-language-unintelligible-to-humans-3876197.html>.

²¹ HAO, Karen. 2019b. Facebook's ad-serving algorithm discriminates by gender and race. *MIT Technology Review*. [En línia] 5 / Abril / 2019b. [Data: 4 / Febrer / 2021.] <https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/>.

currículums. Aquesta automatització del procés de selecció, amb l'objectiu de mecanitzar la recerca del talent, va ser entrenada amb la *dataset* de les candidatures a Amazon dels últims deu anys, la gran majoria homes, el que va comportar que des de la integració d'aquesta tecnologia hi hagués una discriminació a la contractació de dones al penalitzar-se les candidatures d'aquestes.²²

Podem, amb la diferenciació anterior dels orígens del biaix algorítmic, distingir dos tipus de biaix en la IA, biaix cognitiu o biaix generat per la falta de dades. El risc que comporta que les tecnologies més punteres amb IA, fins i tot les *narrow*, puguin incorporar un cert biaix en la seva estructura més bàsica i principal com és la programació de l'algoritme, és un risc que ha de ser observat i regulat pel dret. És necessari que es porti a terme una regulació estricta per evitar la creació de IA esbiaixat degut a la programació de les seves instruccions més bàsiques en l'algoritme. I és que no és cap utopia l'existència de tecnologia IA no esbiaixada, és a dir, sense biaix. La IA pot ser tant o més bona així com ho siguin les dades més primàries i fonamentals del seu propi codi, si es pot obtenir un *dataset* d'aprenentatge completament net de qualsevol prejudici conscient o inconscient, com ara en qüestions de gènere, raça, edat o altres conceptes ideològics, es podrà obtenir un sistema de IA que prengui decisions completament lliures de biaix.²³

Com ja es tractarà de manera concreta més endavant, aquest biaix algorítmic en la programació de qualsevol tecnologia de IA comporta un risc en la utilització dels robots autònoms, risc encara major en aquells amb finalitats militars o malicioses en les que la necessitat de regulació és imperant pels efectes letals o lesius que poden generar.

²² DASTIN, Jeffrey. 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. [En línia] 11 / Octubre / 2018. [Data: 4 / Febrer / 2021.] <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.

²³ KANTARCI, Atakan. 2021. Bias in AI: What it is, Types & Examples, How & Tools to fix it. *AIMultiple*. [En línia] 2 / Febrer / 2021. [Data: 4 / Febrer / 2021.] <https://research.aimultiple.com/ai-bias/>.

3. ASPECTES FONAMENTALS A REGULAR RESPECTE LA IA

Entès el concepte de la tecnologia IA i els seus models i patrons d'aprenentatge és més que destacable la necessitat imperant d'una regulació actual, que pugui no només establir els mètodes necessaris per la programació d'aquesta tecnologia, sinó també el sistema de responsabilitats i obligacions legals a recaure sobre determinats subjectes. I és que si som plenament conscients dels efectes que comporta una nova tecnologia, existeix la capacitat de regular aquests efectes per poder donar-los-hi forma.²⁴ No són poques les propostes regulatòries que s'han presentat en els darrers anys, tant en nivells estatals com internacionals, però cap d'elles obté un fort compromís per part dels actors legislatius. El que tenen en comú totes aquestes propostes és, de fet, la desmitificació de la teòrica neutralitat del sistema de IA, podent realitzar decisions equitatives en el seu procés en pro d'un objectiu.²⁵

3.1. REGULACIONS EN EL DISSENY I DESENVOLUPAMENT DE IA

A nivell internacional, i donada l'afectació que té en el nostre propi territori, són importants les actuacions per part de la Unió Europea, doncs, en la nova etapa legislativa ha centrat un dels seus objectius en la transformació digital, sent fonamental pel Pacte Verd Europeu i per la recuperació de la pandèmia de la COVID-19. És per això que s'ha creat una nova Comissió especial sobre la intel·ligència artificial al Parlament de la Unió Europea. El President d'aquesta nova comissió va expressar la necessitat de desenvolupar una IA que generi confiança, elimini qualsevol tipus de biaix i discriminació i contribueixi al bé

²⁴ MARTÍNEZ QUIRANTE, et al., 2018, *op. cit.*

²⁵ MARTÍNEZ QUIRANTE, et al., 2019, *op. cit.*

comú.²⁶ Seguint aquesta línia descrita pel President de la nova comissió, l'eurodiputat Dragos Tudorache, el Parlament de la Unió Europea ha aprovat tres informes que estudien la regulació de la IA per impulsar i promoure la innovació, la ètica i la confiança en la tecnologia així com la responsabilitat pels danys causats per aquesta.²⁷ La creació d'aquesta nova comissió relativa a la IA és del tot encertada al buscar desfer l'estat actual del mercat en el que no hi ha cap regulació legal pel que fa a la programació ètica i sense biaix d'aquesta tecnologia que cada vegada és més present en tots els sectors i en l'esfera particular dels individus i de la societat en el seu conjunt.

Els intents de crear un marc legal de regulació dels usos i dissenys de les noves tecnologies basades en IA no s'inicien amb la creació d'aquesta comissió, sinó que ja al febrer del 2020 es publica el "LIBRO BLANCO sobre inteligencia artificial – un enfoque europeo orientado a la excelencia i la confianza" en el que s'incorporen els requisits essencials publicats per un grup d'experts al 2019, amb l'objectiu de desenvolupar una IA fiable. La Unió estableix en aquest Libro Blanco una base de set principis essencials sobre la que s'ha de fonamentar tota tecnologia intel·ligent, sent aquests l'acció i supervisió humana, solidesa tècnica i seguretat, gestió de la privacitat i les dades, transparència, diversitat, no discriminació i equitat, benestar social i mediambiental, i la rendició de comptes.²⁸ Amb l'observança d'aquests principis essencials, no vinculants pel moment, es fa palpable la voluntat de la UE de regular i minimitzar les possibilitats reals de risc que comporta la utilització d'aquests sistemes, traduïts en danys materials (per a la seguretat i la salut de les

²⁶ PARLAMENT EUROPEU. 2021. Regulación de la inteligencia artificial en la UE: la propuesta del Parlamento. *NOTICIAS PARLAMENTO EUROPEO*. [En línia] 21 / Gener / 2021. [Data: 4 / Febrer / 2021.]

²⁷ NARRILLOS, Estefanía. 2020. El Parlamento muestra el camino para la normativa sobre inteligencia artificial. *Noticias Parlamento Europeo*. [En línia] 21 / Octubre / 2020. [Data: 4 / Febrer / 2021.] <https://www.europarl.europa.eu/news/es/press-room/20201016IPR89544/el-parlamento-muestra-el-camino-para-la-normativa-sobre-inteligencia-artificial>.

²⁸ COMISSIÓ EUROPEA, 2020, *op. cit.*

persones) com immaterials (dignitat humana, discriminació en l'accés al treball o pèrdua de privacitat).

Un dels altres factors que la UE vol entrar a regular amb força és l'aplicació de les tecnologies de IA tant en els camps militars com en els de justícia i salut, en els que de manera imperativa estableix la base de que la IA és només una eina sobre la que únicament es pren assessorament en la pressa de decisions, que mai podrà reemplaçar o exonerar la responsabilitat humana²⁹. Ressaltant, de manera clara, la necessitat de control humà en totes les decisions preses per qualsevol tipus de tecnologia o sistema de IA, en tots els camps, però en major mesura en el militar, la salut i la justícia.

3.1.1. INCORPORACIÓ DE VALORS I PRINCIPIS ÈTICS EN L'ALGORITME

Un dels problemes que es deriven de la IA no és només el biaix existent en la seva programació més bàsica, sinó que fins i tot en aquells algoritmes totalment lliure de qualsevol tipus de biaix existeix la possibilitat de que el sistema compleixi amb els seus objectius de la manera més competent possible. És a dir, la IA no es fonamenta en la malevolència estrictament, sinó que ho fa sobre la competència per a la qual ha estat dissenyada.³⁰ Aquesta competència pot comportar que la IA utilitzi mitjans totalment contraris als nostres principis més fonamentals per arribar a un objectiu plenament lícit i adequat a aquests principis. Aquesta és una de les situacions que s'han de poder regular de forma més imminent mitjançant el control o supervisió humana en la pressa de decisions com amb altres mètodes de control. Sent necessaris, per evitar l'ús de mitjans incompatibles amb els interessos públics, la

²⁹ YAKIMOVA, Yasmina. 2021. Guidelines for military and non-military use of Artificial Intelligence. *News European Parliament*. [En línia] 20 / Gener / 2021. [Data: 4 / Febrer / 2021.] <https://www.europarl.europa.eu/news/en/press-room/20210114IPR95627/guidelines-for-military-and-non-military-use-of-artificial-intelligence>.

³⁰ O'CARROLL, 2017. *op. cit.*

incorporació d'una regulació que obligui a una visió i incorporació dels valors ètics en el disseny de qualsevol tecnologia d'aquest tipus.

Encara que les regulacions tenen la intenció d'eliminar qualsevol biaix que es pugui donar en aquesta tecnologia no és pas la solució idònia. És a dir, la regulació és necessària encara que no existeixi aquesta discriminació en el sistema de IA, però el que realment s'ha de fomentar és un disseny completament lliure d'aquesta discriminació. Aquest objectiu només es pot assolir, de manera directa i eficaç, en el moment de disseny de la pròpia tecnologia, en les empreses o institucions que es disposin a investigar o invertir en la tecnologia de IA. La percepció de les empreses que treballen amb aquests sistemes segueix aquesta línia, doncs, segons el "2020 State of AI and Machine Learning Report"³¹ només el 15% de les companyies van classificar la diversitat de les fonts i la reducció del biaix com a factors "no importants". Doncs, la impressió general és la necessitat de construir una IA equitativa sense incorporar cap tipus de biaix humà per poder obtenir pressa de decisions i objectius del tot neutrals i realment eficients.

Per poder solucionar el problema des de l'arrel, des de la primera obtenció de dades i des de la primera programació de l'algoritme, és necessari una concepció i visió ètica del treball en aquestes tecnologies, fins al punt de tornar-la imperant mitjançant regulació. La Unió Europea està sent la primera potencia en remarcar la necessitat d'aquesta regulació per poder arribar a ser, fins i tot, un segell característic en la *AI made in Europe*.³² Ja la UE en una de les seves comunicacions sobre la IA³³ tracta el requisits de que les tecnologies han de ser predictibles, responsables, verificables, respectar els drets fonamentals i observar les regles

³¹ APPEN LIMITED. 2020. 2020 State of AI and Machine Learning Report. 2020.

³² COTINO HUESO, Lorenzo. 2019. ÉTICA EN EL DISEÑO PARA EL DESARROLLO DE UNA INTELIGENCIA ARTIFICIAL, ROBÓTICA Y BIG DATA CONFIABLES Y SU UTILIDAD DESDE EL DERECHO. Revista Catalana de Dret Públic. 2019, 58, p. 29-48.

³³ COMISSIÓ EUROPEA. 2018. COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Plan coordinado sobre la inteligencia artificial. Bruselas : s.n., 2018. COM(2018) 795 final.

ètiques. El veritable problema regulatiu deriva de la concepció d'aquests principis ètics al tractar-se d'un terme generalista que pot tenir diverses interpretacions en funció dels factors culturals, ideològics o de sector. De manera internacional s'han arribat a proclamar fins a quaranta-set principis ètics pel projecte AI4People, però els sintetitza en cinc que es consideren que engloben tots ells, via que ha seguit el Parlament Europeu en una de les seves resolucions sobre la matèria.³⁴ Havent de destacar que gran part d'aquest principis ètics són idèntics als principis bioètics, en adaptar-se relativament bé a les noves tecnologies.

Aquest principis, obligatoris en la formulació i disseny de qualsevol tecnologia intel·ligent serien la beneficència, la no maleficència, l'autonomia o acció humana, la justícia i la explicabilitat o transparència. Encara que aquests cinc principis es contemplen en la resolució del Parlament Europeu, aquest també va incorporar l'obligació de respectar els principis de la Carta de Drets Fonamentals de la Unió Europea. El professor Cotino Hueso en el seu article³⁵ realitza una molt adequada, però breu, conceptualització d'aquests principis:

- La beneficència es fonamenta en fer el bé, és a dir, la IA s'ha de desenvolupar pel bé comú i en pro del benefici de la humanitat, però també tenir objectius com la protecció del procés democràtic i l'estat de dret o la sostenibilitat.
- En contraposició es troba el principi de la no maleficència, aquest radica en la prevenció de danys físics, psicològics, financers, socials i al medi ambient. És un principi que limita, doncs, les capacitats futures de la IA i el seu ús fraudulent o indegut, volent evitar, en certa mesura, una carrera bèl·lica en el sector de la IA.
- El principi de justícia no es tracta d'un altre que limitar les discriminacions històriques, evitar el biaix, les estigmatitzacions o altres discriminacions. El

³⁴ PARLAMENT EUROPEU. 2017. Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica. Estrasburg : s.n., 2017. 2015/2103(INL).

³⁵ COTINO HUESO, 2019. *op. cit.*

que es pretén es que la IA distribueixi de manera equitativa els recursos, sense incorporar els prejudicis o les tendències discriminatòries humanes.

- La autonomia o acció humana és un dels primordials per mantenir el control humà sobre les decisions preses per la IA, el que proclama el principi és que encara que amb els sistemes de IA es cedeix part de la decisió a les màquines, aquesta ha de ser reversible, doncs, ha de ser l'humà el que conservi el poder de decidir.
- El principi d'explicabilitat i transparència es podria resumir en la necessitat de poder atribuir responsabilitat a un subjecte determinat, sent necessària la transparència per poder determinar el procediment que ha comportat la presa d'una decisió concreta.

Aquests principis ètics han de ser incorporats a la tecnologia, doncs, no només cal que les persones programadores tinguin una percepció i educació en aquests principis rectors, sinó que s'ha de poder incorporar en el codi junt amb les nocions més bàsiques de l'algoritme, però també realitzant un control *a posteriori* de totes les accions que pren. El que es ve a dir, doncs, és que mitjançant l'aplicació i disseny en base a uns codis de conducta i principis rectors ètics la tecnologia de la IA presenti empatia amb l'ésser humà, no derivada de la capacitat de sentir emocions del propi sistema, sinó que totes les decisions que prengui derivin d'una codificació empàtica i ètica.

3.1.2. CODIS DEONTOLÒGICS PROFESSIONALS EN EL DISSENY I DESENVOLUPAMENT DE IA

En els darrers anys s'ha estat tractant, fins i tot, la possibilitat de crear certs codis ètics o deontològics amb la funció única d'establir els criteris, normes i valors a seguir pels professionals encarregats de dissenyar qualsevol tecnologia intel·ligent amb IA, però també amb les relacions posteriors amb els particulars o clients. Encara existir codis deontològics en sectors com la enginyeria informàtica, un codi

ètic de major amplitud que compregués tots els sectors de professionals implicats en la IA seria, segurament, una manera útil de poder controlar tant *a priori* com *a posteriori* que s'ha aplicat el “bon fer” al tractar-se d'una autoregulació del propi sector. En no existir encara cap codi d'aquestes característiques ja hi ha empreses que han establert el seu propi, doncs, establint les seves pròpies “línies vermelles” pel que fa al desenvolupament de la tecnologia de IA. Un exemple recent és la companyia Bosch auto establint-se el principi rector d'una IA segura, robusta i explicable, mantenint sempre el control les persones sobre aquesta tecnologia.³⁶ Encara que la visió d'aquestes empreses és totalment encertada, encara que només tinguin un criteri empresarial darrera com és l'obtenir la confiança del comprador per incrementar les vendes, comporten una atomització del mercat. Permetre que cada una de les empreses o companyies que dissenyen o realitzen qualsevol altre tasca vinculada amb la IA estableixi els seus propis codis genera una situació de mercat heterogeni que només comporta menys seguretat jurídica al no tenir unes bases comunes a tot el territori.

3.2. REGULACIÓ DE L'ÚS EFECTIU DE LA TECNOLOGIA IA

Arribat el punt en el que, encara tenir una programació del tot ètica i lliure de biaix, el programa realitza una actuació o decisió que infringeixi aquestes pautes, és del tot necessari establir un conjunt de normes que tractin la responsabilitat d'aquestes accions, la identificació del subjecte responsable. Tractant la imposició d'alguna assegurança pels danys o conseqüències futures que pugui comportar l'ús d'aquestes tecnologies de la IA.

³⁶ BOSCH ESPAÑA. 2020. Código ético de AI: Bosch establece las pautas de la compañía para el uso de la inteligencia artificial. *BOSCH Innovación para tu vida*. [En línia] 19 / Febrer / 2020. [Data: 8 / Febrer / 2021.] <https://www.grupo-bosch.es/noticias-e-historias/codigo-etico-de-ai/>.

3.2.1. AUDITORIES I CONTROL EXTERN

Seria interessant, doncs, estudiar la necessitat o obligatorietat de implementar un control de qualsevol tecnologia de IA mitjançant diverses eines, actualment existents, que detecten certs biaixos en els algorismes o en les decisions que pren autònomament el propi sistema. Podem trobar diverses eines i programes, dissenyats per grans companyies tecnològiques, que intenten detectar i solucionar les discriminacions més comunes en tota tecnologia IA: la raça, l'edat, el gènere i la localització.³⁷ Són exemples el Testing With Concept Activation (TCAV) desenvolupat per Google que realitza un anàlisi per detectar biaix en els models de Machine Learning, la mateixa companyia també incorpora el programa What-if Tool que facilita conèixer l'actuació dels models amb les dades del *dataset*. L'empresa IBM ha desenvolupat un dels programes fins a la data més coneguts en aquest camp, l'AI Fairness 360, sent una eina de codi obert que analitza el sistema en busca de qualsevol biaix en la *dataset*, al Machine Learning i al propi algorisme amb l'objectiu de mitigar-lo. Encara que aquestes són les més conegudes i més versàtils en la detecció de discriminacions, hi ha d'altres programes que realitzen les mateixes tasques, com són l'Skater, desenvolupat per Oracle, Audit-AI, desenvolupat per un equip de “*data science*” de Pymetrics³⁸ o la més recent com la de Facebook Fairness Flow.

Aquest control de la IA mitjançant altres tecnologies de IA, encara sonar redundant, no seria més que una auditoria o control extern del sistema intel·ligent. Aquesta auditoria per part de tercers tindria els mateixos objectius que una auditoria financera, per exemple. En comptes d'examinar que es reflecteixi la imatge fidel de

³⁷ AYERS, Ryan. 2020. Tools for Detecting Algorithmic Bias. *Data Science Society, Data.Platform*. [En línia] 30 / Octubre / 2020. [Data: 5 / Febrer / 2021.] <https://www.datasciencesociety.net/tools-for-detecting-algorithmic-bias/>.

³⁸ ASOKAN, Akshaya. 2019. Top 5 Tools Data Scientists Can Use To Mitigate Biases In Algorithms. *Analytics India Magazine*. [En línia] 28 / Maig / 2019. [Data: 5 / Febrer / 2021.] <https://analyticsindiamag.com/top-5-tools-data-scientists-can-use-to-mitigate-biases-in-algorithms/>.

l'empresa, s'examinaria que la tecnologia incorporés totes les regulacions necessàries i que les decisions que es prenen en un simulacre siguin totalment equitatives i encertades als valors ètics establerts. Cal tenir en compte igualment, que en examinar una tecnologia intel·ligent amb un altre mateixa tecnologia es pot donar el problema de que es detecti un biaix pel qual s'ha programat la detecció però no d'altres. Seria considerable plantejar-se el control o auditoria externa del propi sistema de IA que realitza aquest control al no disposar, per ara, d'una tecnologia completament lliure de biaix de qualsevol tipus que pugui realitzar una auditoria totalment fiable.³⁹

3.2.2. RESPONSABILITAT CIVIL

Un dels principals problemes es dona quan una tecnologia, suposadament ètica i neta, que incorpora IA genera qualsevol tipus de dany a l'usuari, tant personal com material. La subsumció de la responsabilitat és un dels punts menys tractats i amb més disparitat en les propostes de regulació. És per això, que en lloc de centrar els esforços en intentar establir una nova legislació, bé a nivell estatal o bé a escala europea, podria ser d'aplicació una de les regulacions actuals que, de manera anàloga, podria aplicar-se a aquestes tecnologies. És cert que es requereix d'una legislació especial sobre la matèria a fi de proporcionar la major seguretat jurídica, però donat que encara no s'han realitzat acords en el sector, és del tot necessari que en els actuals casos en que una IA causi cap dany a una persona es pugui determinar el o els subjectes responsables.

L'aplicació de la llei nacional derivaria de la interpretació anàloga del Reial Decret Legislatiu 1/2007, de 16 de novembre, pel qual s'aprova el text refós de la Llei General per a la Defensa dels Consumidors i Usuaris i altres llei complementaries⁴⁰,

³⁹ COTINO HUESO, 2019. *op. cit.*

⁴⁰ BOLETÍN OFICIAL DEL ESTADO, 2007. Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias. Espanya : s.n., 2007. BOE-A-2007-20555.

però també en compliment de la via proposada per part de la Comissió d'Assumpte Jurídics de la Unió Europea en el seu Informe amb recomanacions destinades a la Comissió sobre un règim de responsabilitat civil en matèria de Intel·ligència Artificial.⁴¹ En aquest Informe es proposa tot un articulat de mesures regulatòries diferenciant per tipus de IA i establint, per exemple, les quanties de les indemnitzacions o diferents règims de responsabilitat.

Centrant-nos, així doncs, en la nostra pròpia legislació al ser l'actualment vigent, incorporant a més certes directives de la Unió Europea, seria d'obligada observació aquest Reial Decret Legislatiu 1/2007, de 16 de novembre. La utilització d'aquesta llei provindria de la definició legal de producte en el seu article 136, sent qualsevol bé moble encara trobar-se unit o incorporat a un altre bé moble o immoble. És a dir, podríem considerar que la tecnologia IA és un bé moble en integrar-se en altres productes o béns, doncs, en l'actualitat està incorporada en *smartphones*, cotxes, maquinaria industrial, electrodomèstics i milers de productes més. El problema que es deriva de l'aplicació anàloga d'aquesta legislació és que no defineix el producte com un bé tecnològic que pot assolir objectius de manera autònoma a través de processos propis intel·ligents que es deriven del Machine o Deep Learning, doncs, processos que poden ser totalment diferents en cada producte, en funció de les dades d'aprenentatge que hagi recopilat. La definició legal de producte defectuós que es fa en la llei (art.137.2) es considerarà com defectuós si no ofereix la seguretat que normalment oferten la resta d'exemplars de la mateixa sèrie. La anàloga aplicació de la llei es deriva, també, d'aquest concepte de producte defectuós. En el cas de la IA encara que totes les incorporades en la mateixa sèrie de productes són iguals en el disseny, cada una d'elles pot arribar a processos diferents determinats per les seves dades d'aprenentatge. Aleshores, no hi ha una seguretat estandarditzada oferta per tota la mateixa sèrie de productes.

⁴¹ COMISSIÓ D'ASSUMPTES JURÍDICS UE. 2020. PROPUESTA DE RESOLUCIÓN DEL PARLAMENTO EUROPEO con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial, de 5 de octubre de 2020.. 2020/2014(INL).

En el Llibre Tercer d'aquest Reial Decret Legislatiu, sobre responsabilitat civil per béns o serveis defectuosos, l'article 128 ens indica que tota persona perjudicada té dret a ser indemnitzat, comprenen la responsabilitat per danys personals, inclosa la mort, i els danys materials sempre que sigui per productes o serveis destinats a un ús o consum privats i hagin estat utilitzats pel propi perjudicat (art.129). Seguint les directrius d'aquesta mateixa llei, el productor serà responsable dels danys causats pels defectes dels productes que fabriquin o importin (art.135), junt amb el proveïdor del producte defectuós que respondrà plenament en cas de que fos coneixedor del defecte, tot podent repetir posteriorment contra el productor (art.146). Un dels punts conflictius amb la interpretació anàloga de la llei seria la càrrega de la prova per part del perjudicat, doncs, en els processos que realitza la IA pot esdevenir, en moltes ocasions, impossible de conèixer per la persona afectada la relació de causalitat o les accions concretes que han portat a una decisió determinada, sense poder determinar la lògica computacional darrere de la decisió. En aquests casos, en estar tractant d'una tecnologia que fa difícil, sinó impossible, la determinació de la causalitat o el fet danyós en concret, seria essencial que es realitzés una inversió de la càrrega de la prova com es fa front a discriminacions per raó de gènere, sent, doncs, la part que discrimina la que provi que la seva actuació o comportament no és discriminatori o danyós.⁴² En les situacions d'incompliment de les condicions contractuals vinculades a la tecnologia, no seria aplicable el criteri de "culpa" com a criteri d'imputació, en ser un mercat en el que la situació d'incompliment prové de la tecnologia IA i no de l'ésser humà que la ha utilitzat, seria idoni crear certs criteris de repartiment del risc de l'incompliment.⁴³

⁴² NAVAS NAVARRO, et al., 2017. *op. cit.*

⁴³ *Ibid.*

3.2.3. ASSEGURANÇA DE LA RESPONSABILITAT CIVIL

La proposta de la Comissió d'Assumptes Jurídics de la Unió Europea⁴⁴ tracta un altre dels aspectes que acompanyen la responsabilitat civil, l'assegurança. No seria difícil plantejar-se la necessitat d'establir una assegurança obligatòria derivat del disseny, venda i, en certs casos, ús de la tecnologia amb IA, ja que en l'actualitat certs sectors o productes ja compten amb aquesta obligatorietat jurídica com a salvaguarda derivada del seu risc. Un simple, però, demostratiu exemple és l'assegurança obligatòria de la circulació de vehicles a motor. En aquesta proposta es defineix la imposició d'una assegurança de la responsabilitat civil obligatòria com a èxit de les noves tecnologies i la confiança de la societat a aquestes encara generar qualsevol dany ocasional. Però també es planteja la problemàtica que comporta la falta de dades sobre els riscos associats a la IA i la incertesa amb la futura evolució de la mateixa, el que comportaria que la Unió Europea col·laborés amb el sector de les assegurances per crear pòlisses que proporcionin una cobertura adequada a un preu no gaire alt. Aquesta voluntat, o necessitat, de col·laboració deriva de la possibilitat de que el mercat privat no pugui esclarir els riscos de la IA i, per tal de garantir la productivitat financera d'aquestes pòlisses esdevinguin a uns preus exorbitants, dificultant o obstaculitzant la investigació i la innovació.

⁴⁴ COMISSIÓ D'ASSUMPTES JURÍDICS UE, 2020. *op. cit.*

4. ROBOTS AUTÒNOMS

Quan tractem termes com sistemes autònoms o IA no només es fa referència a algorismes o programes com els de les xarxes socials o els utilitzats per la banca, sinó que tant el present com el futur de la tecnologia IA està orientat a la robotització. Ens trobem en una revolució industrial iniciada per la incorporació de la robòtica en el sector amb l'objectiu de la cerca continua de la perfecció. Justament, la Indústria 4.0 es caracteritza per la digitalització de les línies de producció i fabricació a través de la realitat virtual, la intel·ligència artificial o la Internet de les coses (IoT).⁴⁵ Encara que aquesta Indústria 4.0 permet aplicar la tecnologia IA en gran quantitat de sistemes robotitzats, són d'especial interès per aquest treball aquells sistemes autònoms que no requereixen d'intervenció humana o quan aquesta és mínima.

La definició de robot, segons imparteixen els professors del Departament de Computació i IA de la Universitat d'Alacant, és una màquina que pot ser programada per fer una varietat de tasques, sent un dispositiu amb la capacitat d'actuar sobre l'entorn i percebre'l.⁴⁶ D'aquesta definició s'extreu que aquells robots autònoms són els que tenen l'aplicabilitat i capacitat de poder executar activitats i tasques sense la necessitat de cap comandament o control directament explícit dels humans. Es caracteritzen, doncs, per poder executar activitats durant llargs períodes de temps sense intervenció humana, amb llibertat de moviment en funció de l'àrea operativa i de terreny, generar rendiment de capacitat d'aprenentatge, realitzar decisions operatives per si mateixos per evitar accions de

⁴⁵ PELEGRÍ, Jordi. 2020. INTELIGENCIA ARTIFICIAL Y ROBÓTICA: LA BÚSQUEDA DE LA PERFECCIÓN EN LA PRODUCCIÓN. *Universal Robots*. [En línia] 11 / Març / 2020. [Data: 14 / Febrer / 2021.] <https://blog.universal-robots.com/es/inteligencia-artificial-y-robotica>.

⁴⁶ CAZORLA QUEVEDO, Miguel Ángel i COLOMINA PARDO, Otto. Robots autónomos. *RUA*. [En línia] [Data: 14 / Febrer / 2021.] <https://rua.ua.es/dspace/bitstream/10045/3547/1/Presentacion.pdf>.

risc altament perjudicials en la seva activitat o arribar a coordinar-se amb altres robots autònoms, tal i com defineixen algunes indústries del sector.⁴⁷

Dels robots autònoms serà interessant tractar de manera més completa els conceptes i graus d'autonomia dels drons i els vehicles autònoms, ja que ens permetrà, posteriorment, vincular-lo amb els robots autònoms letals i les possibles regulacions, a més, són els que actualment ostenten una major representació en el conjunt del sector i la societat.

Els drones, segons el Reial Decret 601/2016, de 2 de desembre, pel que s'aprova el Reglament de la Circulació Aèria Operativa⁴⁸, és un vehicle aeri propulsat que no té personal com a operadors a bord, incloent únicament aquells controlables en els tres eixos. Afegeix les característiques de mantenir-se en vol per mitjans aerodinàmics, sent pilotat de forma remota o amb un programa de vol automàtic, ser reutilitzable i no estar classificat com un blanc aeri, un arma guiada o un dispositiu similar d'un sol ús dissenyada pel llançament d'armes. Aquest tipus d'aeronau es diferencien en civils o d'Estat, sent les últimes les militars amb missió de defensa nacional o enviades per un militar comissionat i les no militars d'ús no comercial, doncs, tota la resta que no obeeix a aquestes característiques es classificada com a civil.⁴⁹ Els graus d'autonomia d'un dron varien en funció del grau de control humà que s'exerceix sobre el sistema algorítmic computeritzat, podent ser autònom únicament en la trajectòria a seguir o prenent totes les decisions durant el vol. La autonomia plena en aquestes aeronaus no tripulades es deriva d'un

⁴⁷ RIPIPSA. 2020. ROBOTS AUTÓNOMOS: QUÉ SON, CÓMO FUNCIONAN Y QUE VENTAJAS OFRECEN. RIPIPSA, S.A. [En línia] 6 / Agost / 2020. [Data: 14 / Febrer / 2021.] <https://ripipsabots.com/robots-autonomos/>.

⁴⁸ BOLETÍN OFICIAL DEL ESTADO, 2016. Real Decreto 601/2016, de 2 de diciembre, por el que se aprueba el Reglamento de la Circulación Aérea Operativa. Madrid : s.n., 2016. BOE-A-2016-11481.

⁴⁹ BOLETÍN OFICIAL DEL ESTADO, 1969. Instrumento de ratificación del Protocolo relativo al texto auténtico trilingüe del Convenio sobre Aviación Civil Internacional (Chicago, 1944), firmado en Buenos Aires el día 24 de septiembre de 1968. (art.3). Madrid : s.n., 1969. BOE-A-1969-1497.

sistema informàtic de IA que interrelaciona el *hardware* amb els sistemes de *software* que controla i dirigeix l'aeronau.⁵⁰

Els vehicles autònoms, són vehicles per definició, però que comporten certes notes característiques en estar equipats amb tecnologies que permeten la conducció sense requerir la forma activa de control o supervisió d'un conductor, segons ha concretat mitjançant instrucció la Dirección General de Tráfico (DGT).⁵¹ En la mateixa instrucció es diferencia que no es considerarà com “autònom” els sistemes de seguretat activa o d'ajuda a la conducció inclosa en els vehicles que requereixen necessàriament control o conducció humana activa. Es podrien classificar dos nivells d'autonomia, els vehicles autònoms i els semiautònoms, seguint l'informe de la Society of Automotive Engineers⁵² en el que es divideixen en sis els nivells d'automatització. Els que es consideren per la instrucció de la DGT com autònoms són els del nivell 3 en el que el conductor actuarà en front a contingències, el nivell 4 en el que certes accions són plenament autònomes però no tota la conducció, i el nivell 5 en el que no cal intervenció humana per cap acció, ni tampoc supervisió. El disseny d'aquests vehicles autònoms és de “*sense-think-act*”, doncs, tot un conjunt de sensors analitzen dades de l'entorn que són processats per un únic processador i interpretats per complexos algorismes que decideixen en fraccions de segons sobre les accions del vehicle per manar als actuadors (frens, llums, columna de direcció, etc.) que les executin.⁵³ Sent necessari que existeixi una interacció entre vehicles (V2V) i amb la infraestructura (V2I) per suplir les limitacions dels sensors.⁵⁴

⁵⁰ NAVAS NAVARRO, et al., 2017. *op. cit.*

⁵¹ DIRECCIÓN GENERAL DE TRÁFICO. 2015. INSTRUCCIÓN 15/V-113, Autorización de pruebas o ensayos de investigación realizados con vehículos de conducción automatizada en vías abiertas al tráfico en general. Madrid : s.n., 2015. INSTRUCCIÓN 15/V-113.

⁵² SOCIETY OF AUTOMOTIVE ENGINEERS. 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. 2018. J3016_201806.

⁵³ NAVAS NAVARRO, et al., 2017. *op. cit.*

⁵⁴ *Ibid.*

5. LETHAL AUTONOMOUS WEAPONS SYSTEMS (LAWS)

La definició de robots autònoms que s'ha realitzat en l'apartat anterior correspon al concepte general, doncs, no diferencia entre robots autònoms civils i militars. En els últims anys la incorporació de tecnologies de IA per desenvolupar robots autònoms amb finalitats bèl·liques no és només una idea de Hollywood. En l'actualitat, tot i no existir una *strong AI* que permeti el desenvolupament d'una robòtica completament autònoma tal i com s'ha estudiat en altres apartats, s'estan realitzant gran nombre d'investigacions en el sector. Existeixen ja sistemes autònoms militars de defensa o *automatic weapons defense Systems* (AWDS), però no són aquests el motiu d'estudi, ja que no emprenen accions letals sinó que serveixen per poder defensar-se front a amenaces en les que la intervenció humana suposaria la diferència entre sobreviure o no. Un exemple és la tecnologia Phalanx, un sistema antimíssils autònoms integrat en molts vaixells militars que permet defensar-se contra míssils supersònics en qüestió de segons, sense necessitat d'intervenció humana.⁵⁵ Doncs, aquests sistemes no seran objecte de tractament en aquest estudi, no només per no ser una tecnologia plenament autònoma en molts casos, sinó per no suposar, pel moment, una amenaça contra la vida humana en no ser sistemes letals. No obstant, mereixen ser objecte de regulació internacional, igual que els LAWS, per poder determinar els seus usos i evitar qualsevol acció contrària de dret o que, fins i tot, els poguessin reconvertir en LAWS tot sent AWDS no dissenyats amb tals finalitats.

La definició de LAWS o *killer robots* és un concepte àmpliament homogeni en la gran majoria dels informes, articles o llibres sobre la matèria realitzats per organitzacions internacionals, científics o experts en IA. Fent referència a l'obra de la professora Martínez Quirante⁵⁶, són sistemes que integren la IA per dotar-los de

⁵⁵ WALSH, Toby. 2020. Killer robots. La intel·ligència artificial i les armes autònomes. 19 / Febrer / 2020, NÚM. 48 INTEL·LIGÈNCIA ARTIFICIAL.

⁵⁶ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

capacitat intrínseca d'abordar processos de decisió al marge del control o supervisió humana de forma significativa, doncs, posseeixen la capacitat d'identificar i eliminar objectius militars, incloses persones, suposant una delegació de capacitats letals a entitats robòtiques. El problema en aquests sistemes autònoms i el risc derivat s'origina en l'autonomia del propi ens robòtic en fases crítiques com són la decisió d'abatre objectius determinats per la IA en base a la seva programació.

Els riscos que es deriva de l'ús d'aquestes tecnologies autònomes i el propi malestar de la societat als usos inadequats ha comportat la generació de diversos comitès, organitzacions o moviments de prohibició legal de l'ús d'aquests sistemes i en la investigació d'aquestes. Però, malauradament, fins al moment no existeix un marc internacional en el que es pugui arribar a un acord en una matèria que cada vegada aixeca més controvèrsia en la societat, en estar més a prop de que aquestes tecnologies puguin utilitzar-se fora de laboratoris o zones de control, doncs, utilitzar-se en el món real.

Els LAWS des de fa uns anys ja venen sent objecte de notícia, i normalment no pels beneficis que se'n puguin generar, sinó pels problemes que han originat o els riscos derivats del seu ús. Són exemples publicacions del 2018, any en el que l'exèrcit dels Estats Units d'Amèrica va anunciar la incorporació de IA en els seus drones MQ-9 Reaper, en les que s'entrava a valorar les fatals conseqüències i la oposició a aquest tipus de tecnologia aplicada a la guerra.⁵⁷ Ja són varis els LAWS amb els que s'està investigant o, fins i tot, que ja estan operatius en diferents exèrcits, sobretot el dels Estats Units d'Amèrica. Són exemples de LAWS terrestres els nous tancs autònoms com el Robot Ripsaw 45 M5 o el Military Robot EMAN dels Estats Units d'Amèrica o el Robot Sorátnik i l'Uran-9 de l'empresa russa Kalashnikov. Però, també s'estan realitzant avenços en LAWS aquàtics com són l'XLUV de la Royal Navy o els de Boeing per l'exercit americà. No només s'estan dissenyant sistemes autònoms robòtics de transport amb armes de foc, sinó que algunes

⁵⁷ LEE, Peter. 2018. Drones will soon decide who to kill. *The conversation*. *Rigor académico, oficio periodístico*. [En línia] 11 / Abril / 2018. [Data: 18 / Febrer / 2021.] <https://theconversation.com/drones-will-soon-decide-who-to-kill-94548>.

companyies/estat van més enllà i desenvolupen robots humanoides amb altes capacitats letals, com és l'Skybot F-850 rus, conegut com a FEDOR.

Encara que es coneixen els riscos que comporten aquests sistemes autònoms letals, que seran estudiats extensament en el següent apartat, les grans potències es centren més en els avantatges militars que comporten. Diversos experts en robòtica i juristes han establert una sèrie d'avantatges tant “suculents” pels Estats com podrien ser la major protecció de les forces armades pròpies al prevenir les baixes de soldats, l'increment de la força utilitzada en el terreny, el comptar amb un temps de reacció molt menor que l'humà o l'ampliació del camp de batalla al poder mantenir-se en el terreny més temps que els soldats humans. Però, sobretot destaquen avantatges econòmiques com són la reducció d'un terç del cost en comparació amb les armes no tripulades actuals i la reducció del cost polític que comporta una guerra.⁵⁸

5.1. RISCOS DELS LAWS

De les pròpies sigles del nom amb que s'ha dotat a aquests sistemes ja es fa present el principal dels perills o riscos que comporta la posta en marxa d'aquestes tecnologies, la vida de les persones. No estem simplement front a discriminacions o errors de programació com en el cas dels algoritmes, sinó que tots els inconvenients que comporta la tecnologia de IA s'amplifiquen en aquests sistemes autònoms.

La Campaign to Stop Killer Robots ha realitzat unes tasques de divulgació i de investigació molt àmplies per poder oferir informes a la població de gran detall i conscienciar, així, sobre la necessitat de prohibició i regulació d'aquests sistemes. Fent referència a les seves publicacions, els problemes en l'ús dels LAWS es

⁵⁸ GUTIÉRREZ ESPADA, Cesáreo i CERVELL HORTAL, María José. 2013. Sistemas de armas autónomas, drones y derecho internacional. 2013, 2 / 2013, p. 27-57.

podrien resumir, de manera més comprensiva per a la població, en sis categories diferents.

1. Aquests sistemes, completament autònoms, podran determinar qui mor i qui viu sense intervenció humana. Sense tenir valors humans com la compassió que són necessaris en la presa de decisions ètiques.
2. Si no es realitza una regulació estricta, les potències que investiguen aquestes tecnologies com són els Estats Units, la Xina, l'Israel, la Corea del Sud, Rússia i el Regne Unit, podrien generar una situació de inestabilitat i cursa armamentística en les armes autònomes.
3. La substitució de persones per robots comportarà que la decisió de declaració de guerra per part dels Estats serà molt més fàcil de prendre, alhora que les accions de robots autònoms de certs estats podrien fer augmentar les tensions en situacions o regions d'inestabilitat.
4. La tecnologia IA no pot realitzar judicis de valor per determinar la proporcionalitat d'un atac o distingir combatents i civils.
5. La incertesa en la subsumpció de la responsabilitat pels actes dels robots plenament autònoms, el que dificultaria la justícia de les possibles víctimes i les responsabilitats internacionals dels Estats.
6. Aquestes tecnologies o sistemes podrien ser utilitzades en situacions no militars, comportant certs riscos per la població civil en situacions de pau, no limitant-se a escenaris de combat.

Aquesta categorització o estructuració en sis grans tipus de risc és molt útil per donar a entendre els punts claus de les implicacions que comporten la utilització d'aquest tipus d'armament. Ara bé, els riscos van més enllà d'aquestes categories.

Humans Rights Watch, planteja en un dels seus molts informes sobre la matèria⁵⁹, una distinció que s'adapta de forma perfecta a la intencionalitat d'aquest estudi. Aquesta organització diferencia entre els riscos legals i aquells perills no legals, a

⁵⁹ HUMAN RIGHTS WATCH. Making the Case. The Dangers of Killer Robots and the Need for a Preemptive Ban. 2016.

més de presentar arguments en contra dels hipotètics, encara, avantatges dels LAWS.

5.1.1. PERILLS NO LEGALS

Seguint la classificació de Humans Rights Watch els perills no legals són aquells que afecten i perceben directament la població, sense tenir en compte les repercussions o regulacions legals en l'ús d'aquestes noves tecnologies. Un d'aquests perills és la pròpia moralitat de les armes autònomes letals ja que, com s'ha demostrat en la secció de la IA, no es pot dissenyar, pel moment, una tecnologia completament ètica i responsable amb un seguit de valors en concordança als socialment acceptats. És a dir, encara que algun dels experts, sobretot els que treballen per les potències involucrades en el disseny d'aquests LAWS, argumenten que es pot arribar a programar aquestes tecnologies en aquest sentit, són molts altres els actors que ja venen expressant les preocupacions en aspectes com la moralitat i ètica dels *killer robots*. I quan ens referim a actors no només es fa referència a científics, el públic en general o premis Nobel, sinó que també han mostrat les seves preocupacions actors tant importants o visibles com són Estats, la ONU o fins i tot la Santa Seu, entre altres líders religiosos. Totes elles centrant-se en la falta de qualitats humanes per garantir la presa de decisions morals i l'amenaça a la dignitat humana que representen els robots autònoms als que se'ls faculta per decidir sobre la vida humana.

Aquesta deficiència de moralitat per la que gran quantitat d'actors mostren la seva oposició no és més que la infracció o amenaça del dret a la dignitat humana i el dret a la vida. S'ha donat en altres ocasions que determinades lleis són declarades inconstitucionals per tractar a les persones com ho farien els LAWS, no com a éssers humans sinó com a objectes o *bytes* d'informació, com va ser el cas de la llei alemanya del 2005 sobre Seguretat del Transport Aeri a Alemanya.⁶⁰

⁶⁰ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

Un altre dels riscos tractat en el mateix informe de Humans Rights Watch és la carència d'emocions humanes que protegirien tant a civils com a combatents. Aquest perill és tractat alhora com una de les grans avantatges de la implementació dels LAWS. Justifiquen l'ús d'aquests sistemes letals amb la creença de que el robot no pot realitzar actes sota les emocions de ira, mal o fam que poden i s'han demostrat, en ocasions, estar relacionades amb atacs indiscriminats a civils. Encara que de certa manera no deixa de ser un avantatge a tenir en compte no s'ha d'oblidar un dels grans riscos que comporta, que segurament no compensa els beneficis que se'n deriven. Les emocions humanes com la empatia o la compassió són sentiments que permeten que els éssers humans siguin reticents a acabar amb la vida d'un altre ésser humà com a igual.⁶¹ És gràcies a aquestes emocions i sentiments humans que en la majoria de conflictes s'han pogut donar situacions com al segle XI amb la prohibició d'atacar a dones, nens i sacerdots o la creació de zones o corredors humanitaris segurs⁶², però que no es podrien dur a terme amb entitats artificials al no comprendre la importància d'aquestes emocions o del propi llenguatge humà no verbal. A més, aquest és un dels perills que més preocupa a la població, no només per evitar baixes en el terreny de combat, sinó que aquesta falta d'emocions podria utilitzar-se per part de moviments totalitaris o populistes per oprimir a tota una societat en temps de pau i geogràficament lluny del conflicte. Ja han hagut declaracions o teories d'aquest estil, com quan Elon Musk es va referir a la superintel·ligència de la IA com a un possible futur dictador immortal del que no ens podríem desfer.⁶³

Un altre dels perills, encara que no té gran justificació, és tot el contrari al que es vol donar entendre amb la resta de perills no legals tractats. És a dir, un dels perills és la pròpia prohibició dels LAWS que comportaria que totes les avantatges que

⁶¹ HUMAN RIGHTS WATCH, 2016. "Making the Case. ..."

⁶² MARTÍNEZ QUIRANTE, et al., 2019. *op. cit.*

⁶³ HOLLEY, Peter. 2018. Elon Musk's nightmarish warning: AI could become 'an immortal dictator from which we would never escape'. *The Washington Post*. [En línia] 6 / Abril / 2018. [Data: 11 / Març / 2021.] <https://www.washingtonpost.com/news/innovations/wp/2018/04/06/elon-musks-nightmarish-warning-ai-could-become-an-immortal-dictator-from-which-we-would-never-escape/>.

aquests poden comportar i la reducció de baixes en les guerres no es puguin dur a terme. Però no és del tot cert, i és per això que no té gaire recorregut aquest risc, ja que actualment sistemes semiautònoms o autònoms no letals, com els ja tractats AWDS, aporten gran part dels avantatges que es vinculen amb els LAWS. Doncs, la única diferència entre LAWS i sistemes no letals semi o autònoms com els que ja estan en el terreny actualment, és la capacitat de que aquests mantinguin els seus objectius fins i tot en situacions en les que les comunicacions són deficientes i sobre la decisió final en moments clau com abatre un objectiu determinat.

5.1.2. *PERILLS LEGALS*

Els perills legals són aquelles situacions o actuacions que en cas de ser realitzades o generades pel sistema LAWS comportaria una infracció de qualsevol tipus de regulació nacional o internacional. Aquests riscos són, sense intenció de treure importància als no legals, una de les principals raons de la necessitat regulatòria o prohibitiva en l'ús d'aquestes tecnologies, doncs, les violacions més comunes de drets són respecte la protecció internacional dels drets humans, al dret internacional humanitari i diversos drets civils. Aquestes vulneracions tan greus de drets si s'analitzen tenint una visió global pel que fa a l'ús dels sistemes autònoms, comporta adonar-se que qualsevol ésser humà de qualsevol país, territori o Estat es pugui veure afectat, ja no només persones totalment alienes al conflicte, sinó persones o poblats que es troben totalment al marge de decisions internacionals o polítiques, posem de cas poblacions indígenes de l'Amazones o altres territoris.

En el dret internacional humanitari existeixen dos principis bàsics o fonamentals, el de distinció i el de proporcionalitat. Aquests principis no són, com podria semblar, meres formalitats per saber quan és aplicable aquesta branca del dret o no. Aquests principis van molt més enllà i tenen el seu origen en els primers conflictes de la humanitat amb una clara influència religiosa, però que no es formalitzen fins

al segle XVI amb els primers ius internacionalistes.⁶⁴ Aquesta branca del dret exigeix que es diferenciïn els objectius legals dels il·legals, és a dir, distingir d'un combatent actiu a un civil o un combatent greument ferit o aquell que ha desertat de les línies enemigues. Els LAWS amb els nivells actuals de sensors i computerització no poden arribar al nivell humà d'identificació d'altres éssers humans mitjançant el llenguatge no verbal, és per això que el primer dels principis ja es veuria vulnerat. Si més no, el principi de proporcionalitat és de vital importància en qualsevol conflicte armat. Aquest es fonamenta en l'estudi o anàlisi de l'equilibri entre el dany civil previst i l'avantatge militar previst. En l'actualitat el criteri emprat per realitzar aquest estudi i esbrinar si un atac o acció militar es proporcional és l'anomenat "comandant militar raonable". És a dir, si un comandant militar raonable hagués autoritzat l'atac en concret, tenint en compte les condicions del moment precís, si l'avantatge militar resultés molt major que el dany civil causat. Aquest criteri d'aplicació per conèixer de la proporcionalitat d'una actuació militar requereix, com és raonable, l'estudi individual del cas i cap dels dissenyadors o programadors no podria registrar en cap algorisme la infinitat de situacions, complexitats o contingències imprevisibles que es poden donar en l'atac per tal que el LAWS estudiï per si mateix la proporcionalitat de l'atac.

Tot arribar a una prohibició d'aquests sistemes no s'estaria exempt de una vulneració del dret internacional. En altres moments de la història ja s'ha donat el cas en que armes condemnades o prohibides per amplitud d'Estats han sigut utilitzades per alguns altres, justificant el seu ús en zones despoblades com ara deserts o el poder ser desactivades a voluntat. En la realitat aquestes justificacions o "promeses" han resultat ser del tot fal·làcies, doncs, en són exemples les municions de dispersió o la reactivació de les mines antipersona per part de Donald Trump al·legant que poden ser desactivades remotament.⁶⁵ Es comprensible que

⁶⁴ JORGE URBINA, Julio. 2000. Protección de las víctimas de los conflictos armados, naciones unidas y derecho internacional humanitario. València : Tirant lo Blanch, 2000.

⁶⁵ EFE. 2020. EE.UU. reintroduce el uso de minas antipersona en sus Fuerzas Armadas. *Agencia EFE*. [En línia] 31 / Gener / 2020. [Data: 11 / Març / 2021.]

una vegada aquests armaments són incorporats a un exèrcit no és, de facto, possible controlar el seu ús per determinar que s'ajustin a les justificacions permeses. Aquests Estats acabaran utilitzant per a determinades situacions aquests sistemes d'armament, però ho classificaran com a secrets d'Estat o de defensa nacional amb l'objectiu de que en un curt termini no es descobreixi que han sigut utilitzats.

Una de les grans preocupacions o perills en l'ús d'aquests sistemes intel·ligents és la determinació de la responsabilitat dels seus actes que vulnerin o causin mal a les persones. Qualsevol ha de poder ser responsabilitzat dels seus actes, tenint el rendiment de comptes un propòsit moral, social, polític i legal, més imperant encara en un sistema emprat a la guerra que pot decidir sobre la vida de les persones. Hi hauria dues branques del dret greument vulnerades, pel que fa al dret internacional humanitari respecte a l'obligació de jutjar els actes criminals que es puguin cometre durant els conflictes armats, i el dret internacional dels drets humans amb l'establiment del principi de reparació per qualsevol abús dels drets humans. Afegint-se la dificultat de considerar a les armes autònomes plenament responsables dels seus actes, els propietaris del sistema o els programadors podrien eximir-se de la responsabilitat des del moment de posada en funcionament del robot i els Estats podrien al·legar que no tenien control sobre el sistema.

Finalment, la vulneració del dret internacional dels drets humans no només és el no poder reparar el mal o dany causat, sinó que un dels principals drets humans, recollits en les Declaracions i Convencions, és el dret a la vida. Aquest dret a la vida en els conflictes armats es defineix com un dret derivat del de dignitat humana, doncs, a no ser assassinat de forma arbitrària, capritxosa, inexplicable, inhumana o com a dany colateral.⁶⁶ I estudiat el concepte de LAWS s'observa una greu vulneració d'aquest dret i d'aquests principis, fonamentalment al trobar-se front a un buit legal.

<https://www.efc.com/efc/america/politica/ee-uu-reintroduce-el-uso-de-minas-antipersona-en-sus-fuerzas-armadas/20000035-4163593>.

⁶⁶ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

6. ASPECTES FONAMENTALS A REGULAR RESPECTE ELS LAWS

6.1. REGULACIÓ PROHIBITIVA

La necessitat imperant d'una regulació és més que evident, doncs, amb tot el llistat de riscos no només legals.

És cert que no estariem front a una matèria que requeriria de regulació en el sentit de desenvolupar un articulat que establís certes limitacions en l'ús d'aquestes noves tecnologies amb finalitat bèl·liques, sinó que ens situariem front a un marc normatiu prohibitiu. Encara que la interpretació del dret internacional humanitari és suficientment mal·leable en permetre destrucció i un ús de la força letal, el potencial destructor es veu amplificat per la simple existència de determinades tecnologies destructives i letals amb capacitats avançades⁶⁷, com són els sistemes autònoms. I és en aquest aspecte en el que tot i complir amb els principis fonamentals del dret internacional humanitari, com és el de proporcionalitat, distinció, necessitat militar i humanitat, entrariem en un dels majors riscos mai experimentats en la història: deixar en mans d'ens robòtics les decisions de jutjar i executar.

Aquesta necessitat de prohibició no només se'n deriva d'aquest risc intrínsec de qualsevol nova tecnologia, sinó que la incorporació d'aquests sistemes armamentístics comporta la modificació de les normes del propi *ius in bello* que tenen per objectiu un enfrontament just en el terreny de combat. Alhora que la influència en les decisions de declarar una guerra són molt majors, plantejant-se la pregunta de que si un Estat sap que no tindrà baixes pròpies en el conflicte serà més irresponsable i tempestiu en la decisió d'atacar.⁶⁸ És per això que és de vital importància que es construeixin barreres, legalment parlant, per poder-nos avançar

⁶⁷ KERR, Ian i SZILAGYI, Katie. 2016. Asleep at the switch? How killer robots become a force multiplier of military necessity. [autor llibre] Ryan CALO, A. Michael FROOMKIN i Ian KERR. *Robot Law*. Cheltenham : Edward Elgar Publishing, 2016.

⁶⁸ *Ibid.*

a les futures conseqüències i perills que la IA porta implícites, més quan s'incorpora en sistemes d'armament autònom que podrien ser comparables a les armes nuclears o sistemes de selecció d'objectius individualitzats.⁶⁹

Una de les principals problemàtiques que recullen alguns autors⁷⁰ és la redacció i inclusió de la precisió algorítmica en la redacció de la llei o regulació corresponent, requerint d'interpretació per part d'aquells encarregats de realitzar tasques de concreció i implementació d'aquestes lleis sobre la legalitat o il·legalitat del possible comportament enjudiciat. D'acord amb altres autors⁷¹ ha de ser interpretat el dret internacional humanitari seguint la doctrina de *technological neutrality* o neutralitat tecnològica. Aquesta doctrina es fonamenta en la concepció contrària d'implementar o interpretar les noves redaccions de les lleis o normatives basant-se en un sector específic o en una tecnologia concreta. És a dir, el dret internacional humanitari té com a fonaments un seguit de principis bàsics que han de ser neutrals respecte qualsevol tecnologia concreta, doncs, les tecnologies militars no són regulades en funció del tipus, sinó quins usos són acceptables pels principis bàsics reguladors d'aquesta branca del dret. Aquesta doctrina de la neutralitat tecnològica permet que les regulacions normatives genèriques siguin eficients amb el pas del temps, doncs, front a la creació o utilització de noves tecnologies, en comptes d'originar-se una nova necessitat regulatòria cada cert temps per ser actualitzada.

Un dels motius legals, ja regulats, que portarien a efectuar una prohibició total d'aquests sistemes autònoms d'armament letals seria la coneguda *Martens clause* o clàusula Martens. Aquesta clàusula està recollida en els Convenis de Ginebra de 1949 i en el seus Protocols Addicionals I i II de 1977, junt al preàmbul del *Convention on Conventional Weapons* (CCW). D'aquesta se'n deriva l'obligació legal per part dels Estats de considerar les implicacions ètiques en la redacció de

⁶⁹ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

⁷⁰ A. SHAY, Lisa, et al. 2016. Confronting automated law enforcement. [autor llibre] Ryan CALO, A. Michael FROMKIN i Ian KERR. *Robot Law*. Cheltenham : Edward Elgar Publishing, 2016.

⁷¹ KERR, et al., 2016. *op. cit.*

noves normatives, sempre que no hi hagi una regulació específica de la matèria.⁷² Alhora que es reconeix la costum com a font del dret humanitari internacional, la clàusula Martens reconeix com a fonts d'aquest dret tant la moralitat en general, com els principis d'humanitat i de consciència pública en particular.⁷³ Arrel de la interpretació de l'anterior doctrina, es determina que segons el dret internacional humanitari s'ha de prohibir el desenvolupament, producció i ús de tot sistema d'armes letals autònoms, doncs aquestes suposen un risc intrínsec de no poder ser utilitzades d'una manera proporcionada i discriminada.⁷⁴ En el cas de poder-se establir certs escenaris en els que sí que actués d'acord a aquests principis, aleshores la clàusula Martens no seria aplicable, sinó que la regulació necessària resultaria una limitació legal del seu ús en aquelles situacions i contextos en els que pogués complir amb les exigències del dret humanitari internacional. És a dir, la diferència respecte la capacitat prohibitiva o no de la clàusula Martens radica en el risc intrínsec d'incompliment o en el potencial incompliment en el seu ús, podent ser prohibida o no, respectivament.

És per tot això necessari una regulació en forma de prohibició efectiva dels LAWS, no únicament respecte a aquestes tecnologies en vulneració de la clàusula Martens, sinó que és igual d'imprescindible la redacció d'una regulació internacional que restringeixi els usos d'aquestes no permetent que es puguin generar connexions entre el desenvolupament de sistemes de defensa o AWDS i els LAWS.⁷⁵ Tot i ser conscients d'aquesta necessitat prohibitòria en el desenvolupament, construcció i ús dels LAWS que es fa pública per part d'operadors jurídics, companyies privades o d'organitzacions nacionals i internacionals, les desavinences entre Estats per acordar una regulació internacional és clara. Encara que en les conferències de la Convenció d'Armes Convencionals (CCW, per les seves sigles en anglès) es

⁷² DOCHERTY, Bonies. 2018. Banning 'Killer Robots': The Legal Obligations of the Martens Clause. *Arms Control Today*. Octubre / 2018, Vol. 48, 8, p. 40-40.

⁷³ ASARO, Peter. 2016. Jus nascendi, robotic weapons and the Martes clause. [autor llibre] Ryan CALO, A. Michael FROOMKIN i Ian Kerr. *Robot law*. s.l. : UK, 2016.

⁷⁴ *Ibid.*

⁷⁵ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

demostra una creixent preocupació respecte a la necessitat de cert control humà en els LAWS, grans potències que ja estan realitzant grans avenços en aquestes tecnologies segueixen oposant-se a una prohibició total. Quedant patents les oposicions a aquesta regulació per part d'Estats com la Xina, l'Israel, Rússia, el Regne Unit o els Estats Units d'Amèrica, entre d'altres Estats que estan iniciant certes inversions en la matèria. No obstant, són 30 els Estats que ja s'han posicionat a favor d'una prohibició total dels LAWS. Però no són suficientment forts per poder iniciar la tramitació adequada d'una regulació internacional que s'espera que tingui caràcter de *hard law*. És per això que un grup d'Estats partidaris d'una regulació forta van recomanar l'any 2018 l'inici de les negociacions sobre la possibilitat d'incorporar certes prohibicions legals per assegurar un control humà significatiu en aquests sistemes autònoms, com a mesura temporal a la contraposició de les grans potències a una prohibició total.⁷⁶

Derivat de l'actual impossibilitat manifesta d'una prohibició total d'aquests sistemes letals autònoms, és necessària, doncs, una regulació que limiti en la majoria del possible els danys que pugui causar aquesta tecnologia i uns principis bàsics rectors del seu desenvolupament, però també respecte del seu ús. Tal com assenyalen els experts, una regulació urgent amb l'objectiu de protegir la investigació d'aplicacions no desitjades, mitjançant sistemes internacionals de control exhaustius i de programes de compliment normatiu en les fases inicials.⁷⁷

Aquesta regulació tindria l'objectiu, temporal fins arribar a una prohibició total, de regular els cinc factors que han de ser identificats abans que un robot pugui legalment fixar i abatre un objectiu.⁷⁸ Derivant cada un d'ells dels preceptes dictats pel dret humanitari internacional, sent:

- i. La responsabilitat de les persones que concedeixen l'autorització de l'ús del LAWS,

⁷⁶ HUMAN RIGHTS WATCH. Stopping Killer Robots. Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control. 2020.

⁷⁷ MARTÍNEZ QUIRANTE, et al., 2019. *op. cit.*

⁷⁸ KERR, et al., 2016. *op. cit.*

- ii. La necessitat militar com a principi del dret humanitari internacional com a criteri d'establiment dels objectius,
- iii. La discriminació de l'objectiu en identificar-lo com a combatent legítimat,
- iv. L'aplicació del principi de doble intencionalitat en definir les tàctiques per establir objectius, aproximació i mantenir-se en una distància segura a l'espera,
- v. I l'aplicació del principi de proporcionalitat en la selecció de patrons de foc.

Totes aquestes regulacions anirien destinades a establir una normativa temporal fins poder arribar al consens polític entre totes les nacions per tal de realitzar una prohibició total dels LAWS. I és que no estaríem front a una novetat en establir un sistema judicial de control exhaustiu i preventiu, doncs, ja des de la creació del Tribunal Penal Internacional totes les activitats o tecnologies potencialment causants de genocidis són sotmeses a aquests controls judicials preventius.⁷⁹

⁷⁹ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

6.2. REGULACIÓ LIMITADORA

Donada la manifesta impossibilitat de realitzar una regulació prohibitiva, sent la imperant en aquest cas, és necessari plantejar una regulació limitadora en la producció, desenvolupament i ús dels LAWS. Aquesta regulació no hauria d'esperar ser una regulació definitiva o a llarg termini, sinó un instrument normatiu que implementés certes limitacions de caràcter vital a tota aquesta activitat relacionada amb els *killer robots* de manera transitòria, orientada cap a la regulació definitiva de prohibició.

És cert, que a diferència de la regulació prohibitiva que hauria de tenir caràcter de *hard law* per totes les implicacions que comporta a nivell internacional, la regulació mitjançant normatives limitadores podria ser orientada com a *soft law*, que encara ser més reduïdes les accions en cas d'incompliment, ajudaria a que el major nombre possible d'Estats s'adherissin a aquesta regulació. Doncs, amb una regulació transitòria de caràcter suau o *soft law* podríem obtenir la participació en aquesta regulació de les majors potències del sector, com són Rússia o els Estats Units d'Amèrica, i possibilitar les converses cap a una regulació del tot prohibitiva.

En aquesta regulació limitadora del desenvolupament, producció i ús dels LAWS s'haurien de contemplar quatre punts o aspectes que són de vital importància per tal de permetre, transitòriament, els robots autònoms letals. Aquests aspectes se'n deriven de les principals vulneracions que els LAWS comporten a les diferents branques del dret, ja sigui nacional o internacional, que van units a les principals preocupacions de la societat. Aquests punts a regular són:

1. La supervisió humana en les fases més crítiques com la d'atacar,
2. Apagat d'emergència front a usos malintencionats o defectuós,
3. No letalitat de les actuacions autònomes, i
4. El règim de responsabilitat en cas de produir-se danys.

6.2.1. LLEIS DE LA ROBÒTICA D'ASIMOV

Abans d'entrar a tractar els aspectes mencionats en la regulació limitadora, és important conèixer les tres lleis de la robòtica proposades per Isaac Asimov al seu llibre "Círculo Vicioso", publicat entre 1941 i 1942. Encara que l'autor les va proposar en un context de ciència ficció, són del tot útils en el desenvolupament de les noves tecnologies de IA, doncs, permeten establir un seguit de normes protectores pels éssers humans.

Asimov va proposar inicialment tres lleis bàsiques, però va acabar introduint la número zero.

0. Un robot no farà mal a la humanitat, o mitjançant la seva inacció permetre que la humanitat prengui mal.
1. Un robot no farà mal a un ésser humà, o mitjançant la seva inacció permetre que un humà prengui mal.
2. Un robot ha de complir les ordres dels humans, excepte quan aquestes ordres contradiguin la primera llei.
3. Un robot ha de protegir la seva pròpia existència sempre que aquesta autoprotecció no posi en conflicte la primera o segona llei.

Aquestes normes es poden vincular de forma adequada als quatre punts clau regulatius proposats per una regulació transitòria limitadora dels LAWS. A més, aquestes lleis d'Asimov estan aplicades en gran part del nostre dia a dia. Són exemples de la primera i segona llei la conducció autònoma de vehicles o d'aeronaus, i de la tercera llei els avisos que alerten als usuaris que la bateria és molt baixa en els smartphones o el retorn a la base de càrrega dels Roomba o robots aspiradors.

6.2.2. SUPERVISIÓ HUMANA

Un dels principals problemes i riscos legals que suposen els LAWS és l'actuació plenament autònoma i sense supervisió o control humà. Partint de la clàusula

Martens, que imposa uns principis d'humanitat i els dictats de la consciència pública en l'avaluació de noves tecnologies armamentístiques, és evident que si es pot regular aquest control o supervisió humana s'estarà donant solució a gran part dels problemes que els LAWS porten aparellats de forma implícita.

Seguint aquests principis d'humanitat que aplica la clàusula Martens, els LAWS denoten una falta d'emoció i de judici legal i ètic. Justament aquesta falta d'emocions i judicis ètics es trobaria en conflicte amb el dret a la vida digna, el respecte a aquesta dignitat, doncs, un tracte humà a les persones.⁸⁰ I és que qualsevol actuació o decisió d'actuar amb força s'ha de prendre i avaluar amb el màxim respecte i dignitat a la vida humana.⁸¹ Donada la falta de sensibilitat i la limitació o nul·la capacitat de poder entendre el concepte de respecte per una vida digna per part de qualsevol robot, tant autònom com no, perilla l'empatia que s'aplica en qualsevol atac sota el control d'éssers humans. L'empatia és una emoció humana, no capaç de ser programada en cap algoritme, que dota a les persones d'un judici ètic i legal per minimitzar els danys i els permet prendre decisions complexes en un context particular. Les màquines letals autònomes no poden exercir aquest judici ètic i legal, sinó que les seves accions estarien fonamentades en els outputs dels algoritmes preprogramats, però que segurament no preveuen tots els factors implicats en la seva programació.⁸² A més, els éssers humans estem dotats, psicològicament, d'una resistència innata a la matança, en ser conscients i realitzar aquest judici sobre la pèrdua de vides. És per això, que una de les necessitats més primordials en la regulació dels LAWS radica en el control humà, només els humans tenim la capacitat de realitzar un judici prudent, és a dir, la capacitat d'aplicar els principis ètics i legals corresponents i determinar en funció

⁸⁰ HUMAN RIGHTS WATCH. Atender la llamada. Un imperativo moral y legal prohibir los robots asesinos. 2018.

⁸¹ HUMAN RIGHTS WATCH. Killer robots and the Concept of Meaningful Human Control. 2016.

⁸² HUMAN RIGHTS WATCH, 2018. "Atender la llamada. ..."

d'experiències anteriors en situacions particulars que pugui determinar el comportament en la present situació.⁸³

Aquestes deficiències que experimentaria qualsevol entitat robòtica preprogramada amb un algoritme capaç de prendre decisions autònomes i letals, limitarien greument la capacitat d'interpretació dels matisos del comportament humà, comprendre la dinàmica política, socioeconòmica i ambiental de la situació. A més, no comprendrien les implicacions o riscos humanitaris de l'ús de la força en contextos particulars, comprometent la capacitat dels LAWS de garantir un tracte humà i de respecte a la vida humana, tant de civils com de combatents.⁸⁴

Per aquest seguit de qüestions és necessari implementar en els robots letals autònoms un control humà, però, una de les majors problemàtiques que es plantegen en la regulació és la proporció de control o el tipus d'aquest control. Destacant que grans potències que investiguen en LAWS, com ara l'Israel o els Estats Units d'Amèrica, no utilitzen els termes de control humà, sinó que prefereixen utilitzar els termes de judici humà apropiat. Aquesta nomenclatura denota la voluntat de no involucrar un control humà, sinó només una certa supervisió dels actes i programar aquest judici humà en l'algoritme del propi sistema.⁸⁵ Tot i així, gran part dels actors a favor d'una regulació dels LAWS advoquen per un control humà significatiu, però en el que possiblement no coincideixen és en el terme imprecís de "significatiu".

Aquesta divergència en el control significatiu és exemplificat per la professora Cummings en un dels seus articles.⁸⁶ Contraposa les situacions en les que es podria entendre que el control és significatiu, però entre elles hi ha grans diferències. La primera seria significatiu que una persona humana activés o iniciés el llançament d'un míssil o disparar amb un arma, estant aquesta a quilòmetres de l'objectiu controlant un drone, o seria també significatiu que una persona humana realitzés un

⁸³ HUMAN RIGHTS WATCH, 2016. "Killer robots and the Concept ..."

⁸⁴ HUMAN RIGHTS WATCH, 2018. "Atender la llamada. ..."

⁸⁵ HUMAN RIGHTS WATCH, 2016. "Killer robots and the Concept ..."

⁸⁶ CUMMINGS, Mary L. 2019. Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification? 5 / Desembre / 2019, Vol. 38, 4, p. 20-26.

seguiment d'un arma fins que aquesta impactés a l'objectiu tenint la possibilitat d'avortar la missió. Entre ambdues situacions hi ha grans similituds com és el control del sistema podent avortar o no atacar a un objectiu, però la gran diferència radica en que en la primera situació és l'humà qui decideix obrir foc, mentre que en el segons dels escenaris l'humà només supervisa que el LAWS obri foc dirigit a l'objectiu correcte. És a dir, hi ha una clara diferència que requereix de regulació sobre aquest control humà significatiu, donat que en ell pot residir el mantenir el dret de respecte a un tracte humà i a la vida digna o la vulneració de tot principi d'humanitat encara mantenir cert control humà.

Aquesta mateixa autora afirma que el control humà significatiu és ple en aquells casos en que es determina l'objectiu a abatre *a priori*, és a dir, abans inclús d'enviar al LAWS en la missió, i en quines condicions seria abatut aquest objectiu fixat. En la decisió o procés de selecció d'objectius existrien dos fases per identificar el blanc i permetre l'acció ofensiva per part de sistemes autònoms. La primera d'elles seria la fase estratègica en la que persones d'alt rang determinen que un objectiu ha de ser abatut per complir amb la finalitat de la missió, sempre complint i respectant els principis que regeixen el dret humanitari internacional, com el de distinció i proporcionalitat. Quan aquestes decisions arriben a escales jeràrquiques inferiors, aquestes persones certifiquen la legitimitat de l'objectiu i les condicions d'aquesta legitimitat. La segona de les fases seria el disseny de l'arma autònoma letal i plantejar a aquesta l'escenari, per tal d'avaluar si aquest sistema autònom podria determinar o establir l'objectiu per sobre de les probabilitat humanes. Encara així, aquest sistema d'armament autònom hauria de ser capaç, tot i identificar els blancs amb major probabilitat que els humans, d'adaptar-se a les condicions del camp de combat per mantenir el respecte a les normes que regeixen l'enfrontament.⁸⁷

En el cas de mantenir aquestes dues fases en el procés de selecció d'objectius per tal de que hi hagi un control o una supervisió humana en les decisions més crítiques de l'actuació dels LAWS no estariem front a un control humà significatiu, sinó més

⁸⁷ *Ibid.*

bé front a una certificació humana significativa.⁸⁸ És a dir, no hi hauria un control humà directe en les decisions del robots autònom o control sobre l'atac a un blanc determinat, sinó que es confiaria en el criteri del sistema autònom fonamentant-se en un seguit de decisions probabilístiques que es determinarien seguint el mètode de doble fase de selecció d'objectius. És, doncs, el sistema pel que advoquen les grans potències com els Estats Units d'Amèrica o l'Israel. Tot i ser cert que amb aquest sistema es manté cert control humà sobre el sistema autònom, no és del tot significatiu. És a dir, sota aquest procés bifàsic en la identificació d'objectius és fixa *a priori* el blanc a abatre, no deixant-lo al lliure arbitri d'un algoritme, però sí es deixa sota el control d'aquest el moment d'atac i la decisió per abatre.

No obstant, aquest sistema de doble fase seria útil per realitzar una regulació de control humà significatiu. Donat que per ara és impossible la regulació prohibitòria dels LAWS, no és cap secret que els Estats que ja investiguen en aquests sistemes seguiran fent-lo, implicant el possible ús d'aquestes tecnologies en situacions de combat real. Si s'estableix una regulació limitadora dels *killer robots* establint un control humà significatiu es podria condicionar o interpretar el sistema de doble fase en favor d'un control humà significatiu i no només com una certificació humana significativa. Aquesta interpretació en favor del control humà implicaria l'obligatorietat de que tot atac o missió que involucri l'ús de LAWS es sotmeti a un control de doble fase en la identificació de l'objectiu a abatre. Doncs, el sistema bifàsic que certes potències busquen es convertiria en un sistema de control obligatori de tota missió. Implicant que aquestes dues fases s'haurien de superar en cada una de les actuacions, havent de ser repetida en cada canvi d'objectiu o de context ambiental, social o cultural que es pugui donar i es pugui conèixer sota el principi de la diligència deguda.

Però aquest sistema de control no seria l'única mesura implementada en busca d'un control humà significatiu, sinó que també seria obligatori, mitjançant regulació, la capacitat de que qualsevol missió sigui avortada per un operador humà. No només es pretén un control humà en el sistema bifàsic de control, sinó que es pugui

⁸⁸ *Ibid.*

mantenir un control de l'arma durant tot el seu ús. Garantint així el control humà significatiu definit per la organització no governamental britànica, Article 36, tenint un control significatiu en les fases més crítiques, sent aquestes el com, on i quan utilitzar un arma, contra què i qui són utilitzades i els efectes del seu ús.⁸⁹

6.2.3. NO LETALITAT I KILL SWITCH BUTTON

Hem d'identificar la principal meta dels sistemes autònoms per adaptar-se al compliment de la llei o "*law-enforcement*" que és l'habilitat d'identificar els subjectes que es consideren objectius de manera correcta i amb grans probabilitats d'encert.⁹⁰ Algoritmes amb grans bases de dades poden obtenir una identificació prou acurada del subjecte, però que com ja s'ha analitzat en els apartats de la IA, la qualitat d'aquesta identificació pot estar esbiaixada o ser de molt poca qualitat. Coneixent aquests riscos associats amb la identificació d'objectius per part de sistemes autònoms, és encara més necessària una regulació que elimini l'ús de força letal de l'equació. En els escenaris en que es produís un fals positiu, és a dir, identificar un civil innocent com a objectiu, per exemple, podria donar-se un atac o lesió contra aquest subjecte que no té relació de cap tipus amb el conflicte, i en canvi, no identificar a un subjecte realment perillós o combatent.⁹¹ És per això, que un dels punts a tractar en la regulació de caire limitador és la introducció del principi de no letalitat i l'obligatorietat d'incorporar certes mesures de desactivació manual d'emergència de tot sistema autònom. Aquestes necessitats es deriven, a més, de la pròpia interpretació de les lleis robòtiques desenvolupades per Asimov, que recordem que no són unes normes programables en l'algoritme del robot, sinó les principals directrius que han de seguir tots els desenvolupadors, fabricants i aquelles persones o Estats que les activen. Els principis de no letalitat i el de

⁸⁹ HUMAN RIGHTS WATCH, 2016. "Killer robots and the Concept ..."

⁹⁰ DARLING, Kate. 2016. Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. [autor llibre] Ian KERR. *Robot Law*. Cheltenham : Edward Elgar Publishing, 2016.

⁹¹ *Ibid.*

desconnexió d'emergència respondria a la primera de les lleis robòtiques, en tenir la finalitat d'evitar la lesió d'un humà, tant de manera directa com indirecta. Doncs, aquesta primera norma no només regula una acció sinó també una omissió, és a dir, la no actuació d'un robot no pot comportar dany, lesió o mort d'un humà.

Ja són varis els moviments legislatius a nivells internacionals que han proposat aquestes necessitats de conferir a les lleis de la robòtica d'Asimov un caràcter normatiu real, és exemple el text aprovat pel Parlament Europeu respecte les normes de Dret civil sobre robòtica.⁹² En aquesta resolució del Parlament Europeu s'estableix tot un marc legal de Dret civil orientat al desenvolupament, construcció i ús de tot tipus de robots. No es refereix concretament a robots armats o letals autònoms, però sí que ha sigut interpretat per la població i certs autors com l'assentament d'unes bases i principis fonamentals en tot ens robòtic, letal o no.

Seguint aquesta línia en el mateix text es tracta la possibilitat de crear una regulació normativa obligatòria en la que s'estableixi la necessitat d'incorporar el que s'ha anomenat en la terminologia anglesa com a “*kill switch*” o botó de desactivat d'emergència. Aquest sistema permetria una desconnexió del sistema autònom, evitant així la producció d'un dany o, en cas de mal funcionament o apropiació indeguda, la producció excessiva de danys.⁹³ Però aquest kill switch no únicament compliria la funció d'evitar la producció d'un dany, sinó que es podria entendre també com la possibilitat de reprogramar el sistema en cas de que l'algoritme no funcionés de forma adequada. La incorporació d'aquests sistemes d'apagat d'emergència no són únics en els LAWS, sinó que es proposen per tota tecnologia que incorpori IA, sent susceptible de poder perdre el control del sistema ràpidament mitjançant el Machine Learning. Resulta necessària la incorporació d'aquests botons d'apagat que permetin una ràpida aturada de qualsevol acció que es trobés

⁹² PARLAMENT EUROPEU, 2017. *op. cit.*

⁹³ KOTTASOVA, Ivana. 2017. Europe calls for mandatory 'kill switches' on robots. *CNN*. [En línia] 12 / Gener / 2017. [Data: 27 / Abril / 2021.] <https://money.cnn.com/2017/01/12/technology/robot-law-killer-switch-taxes/index.html>.

executant el robot que surti de les pautes d'actuació d'aquest.⁹⁴ En els casos dels *killer robots* permetria no només aplicar un control humà significatiu en poder avortar qualsevol missió o atac, sinó que el sistema d'apagat garanteix que qualsevol persona, inclús fora de l'escenari militar, pogués avortar un atac que considerés èticament incorrecte o contrari a dret.

La necessitat d'incorporar aquests sistemes d'emergència esta estretament lligat al principi de no letalitat. Aquest principi no es troba recollit fins al moment en cap proposta regulatòria sobre robòtica, però es podria definir com armes no letals aquelles que es projecten o s'utilitzen amb l'objectiu principal d'inhabilitar a les persones i mitjans materials, minimitzant la probabilitat de causar danys permanents.⁹⁵ L'adopció d'aquest principi podria determinar el compliment del dret internacional humanitari, que junt amb un control humà significatiu implicaria que tots aquells perills o situacions de vulneració de determinats drets universals o humans deixarien de donar-se. Aquest principi de no letalitat, és a dir, la autorització única d'utilitzar els robots autònoms per accions no letals, permetria que la decisió d'abatre un objectiu, humà o material, recaigui sempre sota el control d'un ésser humà, sota unes estrictes ordres d'actuació militars que han de regir-se pel dret en vigor durant el conflicte. Un exemple seria la utilització d'aquests robots autònoms com a mesura per a patrullar deserts o zones de conflicte, però en cas de que l'algoritme identifiqués un objectiu, aquest no seria abatut, sinó únicament retingut o atordit fins a l'arribada d'un comandament humà, els únics per decidir l'acció letal. Aquest principi comporta un problema ja que es dona compliment a la primera de les lleis d'Asimov, compliment que està vinculat a la no aplicació de la tercera llei. Per evitar que el robot pogués executar accions letals o majorment lesives que les proposades seria inevitable que el robot, encara estar sota atac i perillar la seva supervivència, no podria abatre letalment l'objectiu que el seu

⁹⁴ WALCH, Kathleen. 2020. Will There Be A 'Kill Switch' For AI? *Forbes*. [En línia] 5 / Març / 2020. [Data: 27 / Abril / 2021.] <https://www.forbes.com/sites/cognitiveworld/2020/03/05/will-there-be-a-kill-switch-for-ai/?sh=68aa9d952ef5>.

⁹⁵ RODRÍGUEZ-VILLASANTE Y PRIETO, José Luis. 2007. *ARMAS NO LETALES INHABILITANTES Y DERECHO INTERNACIONAL HUMANITARIO*. Bogotá : Cruz Roja, 2007.

algoritme ha establert.⁹⁶ Durant el període entre la “captura” de l’objectiu i la decisió d’aplicar accions letals per part dels humans, permetria que s’analitzés el context, la determinació correcta de l’objectiu i s’establís el càstig oportú, podent ser aquest letal al ser adoptat per persones humanes.

6.2.4. RÈGIM DE RESPONSABILITAT

Un altre dels aspectes primordials a tractar en qualsevol regulació dels LAWS són els règims de responsabilitat. En ser sistemes que incorporen en el seu propi nucli tecnologia de IA en base a algoritmes complexos preprogramats i integrats amb Machine Learning, el règim de responsabilitat podria assimilar-se al que s’ha proposat en el primer bloc d’aquest treball. No s’ha d’oblidar que tots aquells defectes o riscos que comporta la IA, com és el cas del biaix algorítmic, es troben presents en els LAWS d’igual manera, doncs, no deixen de ser entitats robòtiques que funcionen amb tecnologia de IA però amb unes funcions determinades per a situacions o operacions de combat. Encara així, els LAWS suposen més perills que els errors o biaixos de la IA, doncs, requereixen d’un règim de responsabilitat encara més acotat. No obstant, serà aplicable igualment tota la proposta regulatòria tractada en el primer bloc, sobretot durant les etapes de desenvolupament d’aquests sistemes autònoms. Però, donada la gravetat i perillositat de les accions que puguin realitzar una vegada siguin activades per qualsevol organisme, Estat o persona és necessària una regulació, sobretot en les fases d’actuació més crítiques d’aquestes entitats robòtiques letals autònomes.

Un dels majors problemes fins a la data quan es tracta de regular aquest aspecte dels LAWS és que, de facto, és el propi sistema autònom el que identifica i abat un objectiu al seu lliure arbitri. L’excusa de tots els actors involucrats per evitar

⁹⁶ PASQUALE, Frank. 2020. ‘Machines set loose to slaughter’: the dangerous rise of military AI. *The Guardian*. [En línia] 15 / Octubre / 2020. [Data: 27 / Abril / 2021.] <https://www.theguardian.com/news/2020/oct/15/dangerous-rise-of-military-ai-drone-swarm-autonomous-weapons>.

responsabilitats és que des de la posta en funcionament o activació del robot aquests no podien decidir o atacar, per tant, no són responsables.⁹⁷ En la recerca d'aquesta responsabilitat s'han publicat certs informes o resolucions per part d'organitzacions i d'actors legislatius. Un d'aquests actors ha estat el Parlament Europeu en la seva resolució de 16 de febrer de 2017,⁹⁸ recollint com a principi general de les normes de Dret civil sobre robòtica la necessitat de disposar d'una normativa en matèria de responsabilitat, transparència i rendició de comptes que reflecteixi els valors humanistes i universals europeus. A més, reitera en el seu punt Y la necessitat d'establir un règim de responsabilitat civil donat el nivell de desenvolupament de la robòtica i de la IA. S'ha de ressaltar, novament, que aquesta resolució no està orientada a una regulació de la IA aplicada en robots autònoms armats i molt menys letals, però sí que ha estat interpretada com els primers fonaments per regular els LAWS a nivell europeu, tant per una gran part de la població com per la gran majoria de mitjans de comunicació.

Una de les principals problemàtiques en la determinació d'un règim de responsabilitat respecte les actuacions dels LAWS és la gran quantitat d'actors que hi ha involucrats des del desplegament de l'arma fins a aquells que han determinat que actuï en certa manera o pautes.⁹⁹ Podent ser responsables, en constituir actors en alguna de les etapes de desenvolupament, construcció o ús, actors com ara la jerarquia militar i política de l'Estat que ha finançat el projecte i posat en acció el sistema, les persones investigadores o que han avaluat el robot, els programadors informàtics que han generat l'algoritme o el fabricant que l'ha ofert al mercat.¹⁰⁰ La qüestió que es planteja és que cap dels actors que poden tenir certa responsabilitat serà plenament responsable, en no ser decisors directes de les accions que pren de forma autònoma el sistema d'armament.

⁹⁷ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

⁹⁸ PARLAMENT EUROPEU, 2017. *op. cit.*

⁹⁹ TAYLOR, Isaac. 2020. Who Is Responsible for Killer Robots? Autonomous Weapons, Group Agency, and the Military-Industrial Complex. *Journal of Applied Philosophy*. 3 / Novembre / 2020, p. 15.

¹⁰⁰ MARTÍNEZ QUIRANTE, et al., 2018. *op. cit.*

Un altre dels arguments plantejats que porten a no tenir clar el règim a establir és que hi ha dos corrents de pensament respecte la responsabilitat. Tots dos comprenen la necessitat d'un control humà significatiu en totes les fases del cicle de vida d'un LAWS, però els autors de la primera corrent entenen que es pot diferenciar entre el conjunt dels actors que han intervingut i esclarir la responsabilitat en un d'ells. En canvi, la segona corrent reconeix que es pot diferenciar cada un dels actors implicats però que el règim de responsabilitat vindria a ser quelcom un règim de responsabilitat solidària. És a dir, hi hauria una impossibilitat manifesta de determinar una única responsabilitat sobre un subjecte, el que comportaria que es declarés responsable a tot un grup d'individus que han sigut actors en qualsevol de les fases del cicle de vida d'un LAWS.¹⁰¹ Totes aquestes corrents s'aglomeren al debat actual respecte al règim de responsabilitat en les lesions produïdes per robots industrials que incorporen IA, doncs, ja es venen produint accidents mortals a causa dels mal funcionaments d'aquestes tecnologies. En especial, en el cas dels robots industrials amb IA trobaríem una corrent de pensament més, en la que autors com Walker Smith determinen la responsabilitat sobre aquella persona o empresa que ha realitzat la instal·lació i si aquest actor ha realitzat un entrenament correcte del sistema.¹⁰²

Tractant el règim de responsabilitat és més que necessari establir una regulació respecte les possibles afectacions sobre persones o grups de població civil. Com estableixen la majoria d'autors, seguint també les directrius del Dret Humanitari Internacional, és condició fonamental en una guerra "justa" que algú sigui declarat responsable per les morts i lesions de persones durant el transcurs del combat. Sent aquesta responsabilitat més de caire ètic i legal, no tant una responsabilitat causal.¹⁰³

¹⁰¹ TAYLOR, 2020. *op. cit.*

¹⁰² PC PRO. 2015. Killer robots: who's responsible? *PC Pro; London*. [En línia] Octubre / 2015. [Data: 28 / Abril / 2021.] <https://search-proquest-com.are.uab.cat/magazines/killer-robots-whos-responsible/docview/1711126926/se-2?accountid=15292>.

¹⁰³ SPARROW, Robert. 2017. Killer Robots. *Journal of Applied Philosophy*. 1 / Novembre / 2017, Vol. 24.

És necessari, en aquest punt, l'anàlisi dels arguments eximents que serien aplicables als principals actors. Doncs, només així podrem esclarir els punts indispensables a regular. Pel que fa als programadors informàtics de l'algoritme, nucli de la IA, en el cas que el LAWS no actuï de la manera que s'esperaria causant qualsevol acció destructiva, es podria emparar en la bona fe i la no negligència durant les seves tasques de disseny. Els comandaments o personal militar podrien resultar ser responsables en base a la seva participació en el desplegament d'aquestes tecnologies en el terreny, però es declararien exempts des del moment en que el sistema autònom aprèn de les dades que recull i identifica els blancs sense ordres directes dels comandaments. Per tant, no podrien ser plenament responsables els Oficials militars de les accions dels robots i encara menys per les morts produïdes per aquest.¹⁰⁴

Però aquestes eximents han de ser interpretats seguint les doctrines assentades per certs tribunals internacionals. El Tribunal Internacional de Justícia (TIJ) en el cas de 1968 sobre activitats militars i paramilitars dintre i contra Nicaragua, va establir la doctrina en la que declara necessària la prova d'un efectiu control de les operacions militars o paramilitars respecte de les accions que hagin trencat qualsevol disposició del dret internacional per declarar la responsabilitat d'un Estat concret. Més tard, es van concretar les implicacions d'un control efectiu pel TIJ en el cas de 2007 respecte l'aplicació de la Convenció sobre la Prevenció i Càstig del Genocidi. Es va establir que el control és efectiu quan es té control respecte de l'operació específica, no sobre la totalitat de les operacions.¹⁰⁵ Sent del tot correcte aquesta interpretació efectuada pel TJI, ja que presumeix que només un Estat que estigui al corrent i tingui el control de la unitat més petita de les operacions, és a dir, les accions específiques i no només de la generalitat de les accions, pot adequar aquestes actuacions a la disposicions del dret internacional.

¹⁰⁴ *Ibid.*

¹⁰⁵ HUMAN RIGHTS WATCH, 2016. "Killer robots and the Concept ..."

En propostes com la resolució del Parlament Europeu de 2017¹⁰⁶ s'intenta plantejar tot un règim de responsabilitat derivada de les actuacions tant de robots com de tecnologies de IA. D'aquesta en podem extreure una sèrie d'idees que vindrien a ser aplicables al règim específic pels LAWS. Deixant clar, en primer lloc, que qualsevol instrument legislatiu emprat en la regulació d'aquests sistemes d'armament no ha de limitar el tipus o expansió dels danys i perjudicis que puguin ser objecte de compensació, tenint com a única justificació l'autoria no humana. Ara bé, aquesta limitació podria tenir certes excepcions. Tot i que és molt més fàcil aplicar-la en els casos de responsabilitat civil per robots autònoms no letals, no és una idea que pugui quedar desvirtuada en aplicar-se als *killer robots*, si més no. Consistiria en la imposició d'una assegurança obligatòria, d'igual manera que la proposada en el primer bloc per la IA, que permetria que els programadors, fabricants, propietaris o usuaris se'n poguessin beneficiar d'un règim de responsabilitat limitada, juntament amb la contribució a un fons internacional de compensació dels danys i perjudicis ocasionats.

D'aquesta mateixa resolució del Parlament Europeu s'extreu un concepte que suscita vital importància en l'establiment d'un règim de responsabilitat, la imputació de la responsabilitat ha de ser objectiva. Aquesta imputació objectiva es caracteritza per prescindir del requisit de la infracció i no recórrer a la culpabilitat com a criteri d'imputació subjectiva. És a dir, es considera indemnitzable qualsevol dany tot i ser resultat d'un acte permès o legal, i que aquesta imputació és independent de la diligència de l'autor.¹⁰⁷ Doncs, establint un règim d'imputació objectiva només es requereix el provar el dany i l'establiment d'un nexa causal entre el funcionament perjudicial del robot i els danys produïts a la persona o subjectes que els hagin patit.¹⁰⁸ Partint d'aquest criteri d'imputació, resulta obvi que la

¹⁰⁶ PARLAMENT EUROPEU, 2017. *op. cit.*

¹⁰⁷ RUIZ, Enrique i ARAGÓN SÁNCHEZ, Teresa. 2017. Diferencias entre responsabilidad civil subjetiva y objetiva. *LegalToday. Por y para profesionales del Derecho*. [En línia] 12 / Juliol / 2017. [Data: 28 / Abril / 2021.] <https://www.legaltoday.com/practica-juridica/derecho-civil/civil/diferencias-entre-responsabilidad-civil-subjetiva-y-objetiva-2017-07-12/>.

¹⁰⁸ PARLAMENT EUROPEU, 2017. *op. cit.*

responsabilitat ha de recaure sobre l'operador, així com sobre la persona que pugui ordenar l'ús del sistema autònom contra dret. Aquest operador, entenent-lo com totes aquelles parts involucrades tant en el desenvolupament, la construcció i el desplegament d'aquest tipus d'armament, no haurien d'estar sotmesos únicament a un règim d'imputació objectiva. També s'entendria com una obligació de no fer, com s'ha definit en la imputació objectiva, pel que l'eximent de la diligència deguda no resultaria d'aplicació a tot personal que sap o hagi de saber sobre l'experimentació o creació d'aquests sistemes.¹⁰⁹ Aquesta percepció d'operador com un conjunt de persones o subjectes involucrats en les fases de desenvolupament, fabricació i ús dels LAWS respondria a la interpretació de que el Complex Militar-industrial (MIC per les seves sigles en anglès) seria entès com un organisme únic responsable. Per entendre aquesta visió s'ha de conèixer que el MIC està format per la unió entre la indústria armamentística i els serveis de defensa i intel·ligència d'un Estat, doncs, poden diferenciar-se de manera perfecta el govern, l'exèrcit i els dissenyadors dels LAWS.¹¹⁰ Aquesta percepció del MIC com una única organització que esdevindria responsable esdevé de la presumpció en que tots els agents involucrats tenen un interès comú, encara que dificulta la identificació d'un grup d'agents concret plenament responsable permetria la imputació de la responsabilitat en tot el grup.

Resulta necessària la imposició d'un control humà significatiu en totes les fases del cicle de vida de qualsevol tecnologia LAWS, ja que aquest control possibilita l'esclariment de l'origen de qualsevol decisió letal o lesiva produïda per un robot letal autònom, i eliminant així el buit legal respecte de la responsabilitat.¹¹¹ Seguint aquesta tesi en la que amb un control humà significatiu es pot esclarir la responsabilitat, el règim d'imputació objectiva assoleix la seva màxima expressió. Tot i ser una actuació o missió legal, autoritzada pels comandaments militars, aquests comandaments mantindrien en tot moment el control del sistema i podrien prevenir qualsevol atac contrari a dret. A més, amb aquest règim d'imputació, junt

¹⁰⁹ MARTÍNEZ QUIRANTE, et al., 2019. *op. cit.*

¹¹⁰ TAYLOR, 2020. *op. cit.*

¹¹¹ HUMAN RIGHTS WATCH, 2016. "Killer robots and the Concept ..."

amb l'obligatorietat legal del control humà significatiu, es pot fer responsable criminalment al comandament que realitzés un ús autònom o no controlat de l'arma, sense ser determinant si l'acció es jurídica o antijurídica.¹¹² Aquesta imposició d'un règim d'imputació o la imposició d'un règim de responsabilitat penal als robots i als seus creadors podria funcionar com un dispositiu d'autocontrol per les persones que interactuen amb aquests sistemes, però també com a valor instrumental per identificar els subjectes presumptament culpables.¹¹³ Alhora, aquest control humà significatiu permetria determinar la responsabilitat dels comandaments, ja que serien aplicable les doctrines establertes pel Tribunal Penal Internacional i els Tribunals Penals Internacionals de Iugoslàvia i Ruanda. Aquestes doctrines estableixen que si es dona una relació de superior-subordinat, coneixement o amb possibilitat de conèixer de l'acció i la no prevenció del crim o el no sancionar a l'autor, estariem front a la responsabilitat per part del comandament militar. Aquesta doctrina permet, mitjançant el control humà significatiu, que el comandament pugui controlar en tot moment de manera efectiva el LAWS. Seria equivalent a un subordinat més i no deixant la decisió d'actuar amb mitjans letals en mans de l'algoritme. Podent-se evitar la responsabilitat en aquells casos en que el comandament prengué el control de l'arma en aquells escenaris en el que el sistema presentés mal funcionaments o sota l'atac de hackers, per exemple.¹¹⁴

Finalment, aquesta necessitat d'imputació objectiva de la responsabilitat es desvirtua, encara tenir una normativa del tot concreta, si no es segueixen dos principis fonamentals. Aquests principis ja venen d'alguna manera esmentats i intentats desenvolupar en la resolució del Parlament Europeu¹¹⁵ en els seus apartats 12 i 59.e). Es tracta d'establir el principi de transparència que possibiliti sempre la justificació adoptada per part del sistema autònom de manera comprensible pels humans i que pugui tenir impacte significatiu sobre la vida d'una o varies persones, en el cas dels LAWS sempre. Aquesta transparència involucra la inclusió de certs

¹¹² HUMAN RIGHTS WATCH, 2016. "Killer robots and the Concept ..."

¹¹³ MARTÍNEZ QUIRANTE, et al., 2019. *op. cit.*

¹¹⁴ HUMAN RIGHTS WATCH, 2016. "Killer robots and the Concept ..."

¹¹⁵ PARLAMENT EUROPEU, 2017. *op. cit.*

sistemes de registre, similars a les caixes negres en l'aviació, per tal de realitzar un registre de totes les dades que ha utilitzat el sistema autònom per arribar a emprendre certes actuacions. Aquest principi buscaria el poder estudiar en els casos de mal funcionament, danys a civils o descontrol del LAWS, el registre de dades que ha portat a aquesta actuació inapropiada del robot armat autònom i poder determinar, si és el cas, la responsabilitat d'un grup d'actors plenament responsable, per possible negligències o errors bàsics de programació, per exemple, o la responsabilitat del MIC al complert.

L'altre principi completament lligat a la responsabilitat, possiblement de vital importància per poder determinar-la, és la creació d'un sistema d'identificació del propi robot. És a dir, esdevindria un registre d'identificació similar al que s'utilitza en les armes actuals però amb les funcionalitats del sistema de matrícules dels automòbils o els símbols nacionals en els avions de combat. La creació d'un número de matrícula amb el seu corresponent registre internacional asseguraria l'associació entre el robot i la organització a la que depèn, evitant que es puguin donar situacions en les que els grups o persones perjudicades no poguessin denunciar els danys o lesions que han patit en no saber o conèixer la propietat dels sistemes autònoms. Aquest sistema d'identificació evitaria en major mesura la perpetuació de certs crims de guerra, doncs, seria completament contrari a dret el desplegament d'aquest tipus d'armament sense les identificacions concretes i permetent que aquestes realitzessin actuacions contraries al dret internacional humanitari. Actuaría com un instrument preventiu per evitar que els Estats, sota la percepció d'impunitat, realitzessin determinades actuacions, ja que amb aquest registre i sistema d'identificació és podria realitzar de manera senzilla la traçabilitat del LAWS.¹¹⁶

¹¹⁶ PASQUALE, 2020. *op. cit.*

7. CONCLUSIÓ

Del propi concepte de IA es deriva tota una sèrie de riscos que podrien, si més no, perjudicar a la imatge general de tota aquesta tecnologia. La IA encara presentar grans i diversos riscos, tant personals com materials, genera una molt major quantitat de beneficis o avantatges, fins i tot en el simple dia a dia de qualsevol individu.

La regulació d'aquest sector no només obeeix a les necessitats de prevenir qualsevol dany, sinó que també respon a la exigència de garantia i confiança que la societat demana a qualsevol producte. Seguint les pròpies disposicions de l'actual legislació, tota persona té dret a no ser objecte d'una decisió fonamentada, únicament, en el tractament automatitzat o en l'elaboració de perfils que puguin produir efectes jurídics en el propi individu (art.22.1 RGPD). Aquesta simple idea deriva dels riscos que comporten circumstàncies com el biaix en els propis algoritmes, que fa que les tecnologies de IA no siguin del tot equitatives com és d'esperar. Aquest biaix cal regular-lo, disminuir-lo i, finalment, eliminar-lo des del propi moment en el que qualsevol companyia o institució es decideixi per investigar en aquestes tecnologies. I com s'ha indicat en els propis apartats, no és possible obtenir una IA lliure de tota discriminació sense la incorporació d'uns valors i principis ètics, tant en el personal investigador com en la pròpia codificació del sistema. Encara que els programadors poden introduir milers de paraules ofensives en el sistema per donar-los-hi què buscar, el sistema segueix sense poder comprendre el context per esclarir si qualsevol altre paraula és ofensiva. Sent el mateix, la tecnologia amplifica, com ja s'ha demostrat, tots aquells biaixos presents en l'ésser humà. Fins i tot, dotant a la tecnologia de totes les eines per realitzar correctament la recerca de la neutralitat, la no comprensió del context o de l'entorn pot comportar que aquesta realitzi una acció discriminatòria sense etiquetar-la com a tal.

Però no només es requereix d'una regulació en el desenvolupament o investigació de la IA, sinó que cal establir tota una sèrie de règims de responsabilitat o

subsumpció de la responsabilitat civil pels possibles danys que es derivin del seu ús. La Unió Europea ens proporciona una proposta marc que no va gens desencaminada, en quant al règim indemnitzador com amb la subsumpció. Tot i així, cal un veritable consens polític per poder dotar de vinculació efectiva aquestes disposicions, i per ara no el hi ha. No s'ha de descuidar que encara no existir un marc regulador específic plenament vinculant, si que s'ha de complir tota aquella legislació que ho és, doncs, qualsevol tecnologia ha de respectar i vetllar la garantia dels principis i drets fonamentals que caracteritzen a la pròpia Unió Europea i dels que dota a tots els seus ciutadans, recollint-los en diversos textos legals.

És totalment imprescindible el realitzar auditories internes i externes vetllant pel correcte compliment de tots els requisits en l'ús d'aquestes tecnologies i per detectar aquelles actuacions malicioses o defraudatòries que puguin comportar un major risc del legal i socialment acceptat. No és res estrany l'ús d'auditories amb la intenció de complir aquests mateixos objectius, des de fa anys les empreses amb certes característiques com la dimensió o facturació han de suportar aquesta càrrega per esclarir si tant els comptes com les accions públicament presentades responen a la imatge fidel de l'empresa o organització. Buscant-se el principi bàsic de la transparència amb el que generar una confiança de la societat a una certa activitat o sector. Aquestes auditories no s'haurien de deixar de manera única al mercat privat, sinó que seria convenient la creació d'un òrgan comunitari o internacional especialitzat que realitzés aquesta mateixa tasca, per dotar d'una major confiança al sector.

Els problemes actuals, com la falta de consens o interessos diversos i contradictoris entre autoritats i companyies, ha conduït a una situació de molt limitada regulació o, fins i tot, de mercat anòmic. Però, no oblidem que és obligació dels governs i del sector públic, en tant que responsabilitat amb els ciutadans, de negociar amb les grans companyies i investigadors tecnològics per prevenir tota classe de biaix, abusos de poder, responsabilitat per danys, règims indemnitzadors i obligacions d'assegurança. Destacant la visió d'un gran conjunt d'empreses que han observat i tractat aquesta situació, però que malauradament, tot i ser un pas al front, fins comptar amb una regulació legal no fan més que atomitzar i dividir el mercat,

generant més confusió o malestar per part dels usuaris. Aquest mandat cap a les autoritats respon a la transferència social de prejudicis de certs grups a la IA, formant-se un cercle viciós del que serà difícil sortir i que podria arribar a comportar que no sigui la tecnologia o els algoritmes els que s'adaptin a nosaltres, sinó a l'inrevés. És per això, que en qualsevol dels marcs regulatoris que es pugui acabar implementant, és d'obligada redacció la incorporació d'un "interruptor d'apagat" en qualsevol de les tecnologies de IA per poder exercir el control humà en qualsevol situació en que arribés a ser necessari. Es perseguiria així el compliment del principi de reversibilitat i resiliència per poder apagar o desconnectar qualsevol tecnologia que actués de manera contrària als valors ètics o altres regulacions, fins i tot de les tecnologies que realitzen el control o auditoria d'altres tecnologies IA.

Tot i que el camí fins arribar-hi és llarg i complex, és totalment urgent iniciar aquestes negociacions per poder donar forma a les regulacions futures en un sector que cada dia està més present en qualsevol societat, i que cada vegada més serem testimonis dels riscos que comporten.

Pel que fa als LAWS, en aquest sector tan complex i nou en incorporar tecnologia de IA en sistemes d'armament, permetent que siguin totalment autònoms, es generen uns riscos als que la humanitat no ha fet front mai abans. Riscos molt més alts que els beneficis que pot aportar. És aquest escenari d'escalada en el desenvolupament i disseny de nous sistemes armamentístics amb tecnologia IA integrada, el que ens determina la necessitat d'establir el més aviat possible una regulació de caire prohibitòria de qualsevol investigació, i molt més en l'ús, d'aquests sistemes autònoms que no requereixen d'interacció humana. Ara bé, estudiada la dificultat o impossibilitat actual d'arribar a aquesta regulació, no es justificable el no intentar fer petites passes endavant. És per això que s'ha recollit en els apartats del segon bloc del treball tot un seguit de disposicions que, donades les implicacions que els LAWS tenen sobre la vulneració de les branques del dret més fonamentals i internacionals, entrarien a regular de forma inicial i bàsica els punts més conflictius. La simple investigació en aquests sistemes autònoms, encara més el desplegament d'aquestes armes en el terreny, esdevenen un gran risc per a

la població global, junt amb les vulneracions que comportarien del dret internacional humanitari en cas de ser emprades en conflictes armats.

Tota aquesta regulació que es proposa no és amb l'únic objectiu de limitar o prohibir aquesta tecnologia en el sector concret de la indústria armamentística, sinó que busca, també, evitar els problemes associats que la militarització de la IA pot originar, com són el propi futur d'aquesta tecnologia en qualsevol altre sector no militar en generar-se una percepció de la societat en contra d'aquesta tecnologia, com va succeir amb la nuclear o la química. Aquesta urgència regulatòria sorgeix del propi fet que com més es desenvolupi la tecnologia de IA i quant més autònoms siguin els robots més difícil resultarà considerar-los simple instruments en mans de determinats agents. A més, amb aquesta pèrdua d'instrumentalitat humana en els robots trobem que en l'actual marc jurídic els robots no poden ser considerats responsables dels actes o omissions que causen danys a tercers, pel que es requereixen normes i principis específics que aportin claredat sobre la responsabilitat pels actes i omissions dels robots que no es puguin atribuir a un agent humà concret.

És per tot això que en el cas de no poder-se realitzar una prohibició i haver de permetre cert ús de sistemes d'armament autònom o LAWS, la comunitat internacional ha d'obligar a que pel seu ús s'assegurin uns nivells de certificació humana del tot estrictes, tant en la fase estratègica de selecció d'objectius com en la de disseny del propi sistema. Havent de ser provades i requerides a superar uns nivells d'actuació durant entrenaments sota criteris objectius i rigorosos, demostrant així una millor actuació o probabilitat d'actuació que la d'un humà front a una situació idèntica. Però no només s'hauria de regular aquest aspecte tan concret, sinó que hauria de ser tractat com un deure dels Estats el que aquestes armes no puguin lesionar o violentar els drets humans tant en situacions de conflictes armats com en situacions de pau.

Per poder garantir totes aquestes actuacions del robot autònom d'acord al dret internacional és imprescindible que es regulin de forma immediata aspectes claus com són el control humà significatiu. Com ja s'ha tractat, aquest control no implica

únicament el posar en mans d'un militar l'acció de control un botó vermell per atacar quan aquest s'encengui, sinó que el control humà significatiu requereix de control i supervisió humana durant totes les fases al voltant del cicle de vida del sistema autònom. Aquestes fases impliquen un control humà des del moment en que un equip científic o gabinet militar decideix posar en marxa un projecte d'investigació en el que acabar desenvolupant un sistema robòtic autònom letal. És per això que en tota aquesta regulació corresponent al control humà significatiu des del moment de disseny de l'algoritme de la IA s'ha d'aplicar la regulació relativa a evitar el biaix algorítmic que es tracta en tot el primer bloc d'aquest treball.

Encara així, el biaix algorítmic i els errors o programacions amb usos fraudulents es poden seguir donant, per molta regulació existent. És el que s'acostuma a conèixer popularment com “feta la llei, feta la trampa”. Però com tota regulació, ha d'establir un seguit de normativa i disposicions amb l'objectiu d'establir un règim sancionador i de responsabilitat objectiva que busqui el poder esclarir la responsabilitat i les implicacions que aquests usos tinguin en cas de produir qualsevol dany. Per poder establir tota aquesta regulació, el que finalment marcarà la diferència serà la disposició o no de control humà significatiu en totes les fases pel procés de l'acció letal. Sense aquest control humà significatiu resulta pràcticament impossible l'intent de creació d'un règim d'imputació, i és que en cas de víctimes civils és imperant que el responsable pugui ser declarat com a tal, sent aquest un principi bàsic i contundent del Dret Internacional Humanitari. Seguint aquest requeriment de poder establir la responsabilitat de l'actor, ja sigui el del programador, del fabricant o del comandament militar, és fa més necessari en entendre que la dignitat humana es veuria greument afectada en aquelles situacions en que la víctima intenta apel·lar a la humanitat de l'executor i es trobes que aquesta és denegada per estar tractant amb un ésser artificial.

Aquesta apel·lació a la dignitat humana, converteix en essencial interpretar els escenaris de combat amb una mínima expressió de respecte a la moralitat i a l'enemic, és a dir, si deixem la capacitat i la decisió de prendre la vida a un ésser humà en mans d'entitats robòtiques autònomes aquesta decisió es realitzaria sense cap indicatiu de moralitat. Aquesta s'obté quan algú pot ser declarat responsable per

les morts que causa, i poder informar als seus familiars el motiu pel qual ha mort, el que inclou conèixer el responsable i les justificacions del perquè ha sigut mort en combat. Sent, a més, aquest requeriment de conèixer i poder determinar un actor com a responsable una obligació en el *ius in bello*. En aquest aspecte es podria realitzar certa comparació amb les armes de destrucció massiva o les mines antipersona, doncs, en totes dues no es compleix el principi del Dret Internacional Humanitari de distinció. En el cas de les mines antipersona vulneren encara més aquest principi, però també el poder esclarir el responsable de la lesió, doncs, no hi ha una imputació de la responsabilitat en no existir un actor que decideixi atacar o no a aquell civil, que per accident ha trepitjat un d'aquests artefactes. Mostrant-se en la utilització d'aquestes armes una gran i greu falta de respecte cap a la vida humana de qualsevol persona.

I és que el veritable problema resideix en que, tot i arribar a un algorisme perfecte, qualsevol procés autònom dissenyat o programat amb anterioritat, basat en dades recollides per sensors, no realitza cap judici legal i tampoc moral. Tenint en compte que aquells actors que tinguin el control d'aquests sistemes autònoms, control en el sentit de poder comprar aquest tipus de tecnologia, aplicaran un gran poder sobre la població, tant en conflictes armats com no, afectant fins i tot a aquells sectors socials que acostumaven a tenir certs tractes privilegiats i en situacions no estrictament de combat. Perillant encara més la vida de qualsevol persona sense importar la seva procedència amb la capacitat de la IA de poder generar certs models predictius. Amb aquesta capacitat el sistema autònom podria “determinar” si un subjecte segons les seves accions o manifestacions podria convertir-se en un futur objectiu. Doncs, el LAWS actuaria com a un jutge i executor preventiu, en casos en els que determinés que per les formes d'expressar-se d'un individu es mostren indicis de radicalització futura, per exemple.

Tot aquest seguit de propostes regulatòries té la finalitat de poder realitzar una transició de sistemes autònoms letals passant a ser més properes a forces de manteniment de la pau o amb tasques de vigilància, integrant el principi de la no letalitat ja descrit, per exemple. Buscant el manteniment de la doctrina de la sentència del Tribunal Militar Internacional de Nuremberg contra els criminals de

guerra alemanys de l'1 d'octubre de 1946, en la que s'afirma que els crims son comesos per homes i no per entitats abstractes, i únicament amb el càstig dels responsables es poden fer complir les disposicions del dret internacional. Que en el cas dels LAWS, per poder garantir aquesta responsabilitat és del tot necessari poder obtenir una traçabilitat del propietari mitjançant un sistema d'identificació tant del creador, controlador o propietari amb el que es podria la víctima dirigir contra el que l'ha lesionat o atacat, assegurant així la responsabilitat d'aquest. Establint tot un règim sancionador en cas de no complir amb aquesta mesura, doncs, l'incompliment portaria a la vulneració de tota la resta de la normativa.

És per tot això que la comunitat internacional, sent conscient del risc que comporta, ha iniciat una sèrie de debats i cimeres per arribar a una regulació que compleixi amb els principis legals de precaució, distinció i proporcionalitat. Demostrant-se aquest interès amb les discussions i convencions portades a terme per les Nacions Unides, que junt amb la Convenció d'Armes Convencionals (CCW) han tractat de forma extensa el requeriment de deixar l'última decisió respecte la vida d'una persona en mans i control d'un humà o d'una màquina. Sent destacable que en sectors dintre de les pròpies branques militars de certs Estats ja han manifestat, de manera tàcita, estar en contra d'aquests sistemes autònoms letals. Rebuig provinent de l'afectació pròpia que tenen certes armes en els propis soldats. Determinats artefactes el gas taronja utilitzat a Vietnam, no només varen afectar-los directament, sinó també a les seves futures generacions. I el risc de que "miniLAWS" puguin arribar a desenvolupar-se, durant el transcurs d'una efectiva regulació, i recaure en mans de civils o persones amb intencions malicioses segueix creixent dia a dia.

BIBLIOGRAFIA

A. SHAY, Lisa, et al. 2016. Confronting automated law enforcement. [autor llibre] Ryan CALO, A. Michael FROOMKIN i Ian KERR. *Robot Law*. Cheltenham : Edward Elgar Publishing, 2016.

ACTIONS DATA. 2020. Big Data: Cómo lo usan las redes sociales. *ActionsData*. [En línia] 30 / Gener / 2020. [Data: 3 / Febrer / 2021.] <https://www.actionsdata.com/blog/big-data-como-lo-usan-las-redes-sociales>.

APPEN LIMITED. 2020. *2020 State of AI and Machine Learning Report*. 2020.

ASARO, Peter. 2016. Jus nascendi, robotic weapons and the Martes clause. [autor llibre] Ryan CALO, A. Michael FROOMKIN i Ian Kerr. *Robot law*. s.l. : UK, 2016.

ASOKAN, Akshaya. 2019. Top 5 Tools Data Scientists Can Use To Mitigate Biases In Algorithms. *Analytics India Magazine*. [En línia] 28 / Maig / 2019. [Data: 5 / Febrer / 2021.] <https://analyticsindiamag.com/top-5-tools-data-scientists-can-use-to-mitigate-biases-in-algorithms/>.

AYERS, Ryan. 2020. Tools for Detecting Algorithmic Bias. *Data Science Society, Data.Platform*. [En línia] 30 / Octubre / 2020. [Data: 5 / Febrer / 2021.] <https://www.datasciencesociety.net/tools-for-detecting-algorithmic-bias/>.

BBVA COMMUNICATIONS. 2017. El poder predictivo de las redes sociales. *BBVA*. [En línia] 9 / Agost / 2017. [Data: 3 / Febrer / 2021.] <https://www.bbva.com/es/poder-predictivo-redes-sociales/>.

BOLETÍN OFICIAL DEL ESTADO. 1969. *Instrumento de ratificación del Protocolo relativo al texto auténtico trilingüe del Convenio sobre Aviación Civil Internacional (Chicago, 1944), firmado en Buenos Aires el día 24 de septiembre de 1968. (art.3)*. Madrid : s.n., 1969. BOE-A-1969-1497.

—. **2016.** *Real Decreto 601/2016, de 2 de diciembre, por el que se aprueba el Reglamento de la Circulación Aérea Operativa.* Madrid : s.n., 2016. BOE-A-2016-11481.

—. **2007.** *Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias.* España : s.n., 2007. BOE-A-2007-20555.

BOSCH ESPAÑA. 2020. Código ético de AI: Bosch establece las pautas de la compañía para el uso de la inteligencia artificial. *BOSCH Innovación para tu vida.* [En línea] 19 / Febrer / 2020. [Data: 8 / Febrer / 2021.] <https://www.grupo-bosch.es/noticias-e-historias/codigo-etico-de-ai/>.

CAZORLA QUEVEDO, Miguel Ángel i COLOMINA PARDO, Otto. Robots autónomos. *RUA.* [En línea] [Data: 14 / Febrer / 2021.] <https://rua.ua.es/dspace/bitstream/10045/3547/1/Presentacion.pdf>.

COMISSIÓ D'ASSUMPTES JURÍDICS UE. 2020. *PROPUESTA DE RESOLUCIÓN DEL PARLAMENTO EUROPEO con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial, de 5 de octubre de 2020.* 2020. 2020/2014(INL).

COMISSIÓ EUROPEA. 2018. *COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Plan coordinado sobre la inteligencia artificial.* Bruselas : s.n., 2018. COM(2018) 795 final.

—. **2020.** *Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza.* Bruselas : s.n., 2020. COM(2020) 65 final.

COTINO HUESO, Lorenzo. 2019. *ÉTICA EN EL DISEÑO PARA EL DESARROLLO DE UNA INTELIGENCIA ARTIFICIAL, ROBÓTICA Y BIG*

DATA CONFIABLES Y SU UTILIDAD DESDE EL DERECHO. *Revista Catalana de Dret Públic*. 2019, 58, p. 29-48.

CUMMINGS, Mary L. 2019. Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification? 5 / Desembre / 2019, Vol. 38, 4, p. 20-26.

DARLING, Kate. 2016. Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. [autor llibre] Ian KERR. *Robot Law*. Cheltenham : Edward Elgar Publishing, 2016.

DASTIN, Jeffrey. 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. [En línia] 11 / Octubre / 2018. [Data: 4 / Febrer / 2021.] <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.

DIRECCIÓN GENERAL DE TRÁFICO. 2015. *INSTRUCCIÓN 15/V-113, Autorización de pruebas o ensayos de investigación realizados con vehículos de conducción automatizada en vías abiertas al tráfico en general*. Madrid : s.n., 2015. INSTRUCCIÓN 15/V-113.

DOCHERTY, Bonies. 2018. Banning ‘Killer Robots’: The Legal Obligations of the Martens Clause. *Arms Control Today*. Octubre / 2018, Vol. 48, 8, p. 40-40.

EFE. 2020. EE.UU. reintroduce el uso de minas antipersona en sus Fuerzas Armadas. *Agencia EFE*. [En línia] 31 / Gener / 2020. [Data: 11 / Març / 2021.] <https://www.efc.com/efe/america/politica/ee-uu-reintroduce-el-uso-de-minas-antipersona-en-sus-fuerzas-armadas/20000035-4163593>.

FRABETTI, CARLO. 2017. Las leyes de la robótica ¿Cumplirán los robots reales, que ya están entre nosotros, las tres leyes de la robótica de Asimov? *El País*. [En línia] 25 / Agost / 2017. [Data: 2021 / Març / 24.] https://elpais.com/elpais/2017/08/24/ciencia/1503574908_187790.html.

GONZÁLEZ, Andrés. 2019. ¿Qué es Machine Learning? *Cleverdata*. [En línea] 2019. [Data: 3 / Gener / 2021.] <https://cleverdata.io/que-es-machine-learning-big-data/>.

GUTIÉRREZ ESPADA, Cesáreo i CERVELL HORTAL, María José. 2013. Sistemas de armas autónomas, drones y derecho internacional. 2013, 2 / 2013, p. 27-57.

HAO, Karen. 2019. Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo. *MIT Technology Review*. [En línea] Technology Review Inc., 8 / Febrer / 2019. [Data: 3 / Febrer / 2021.] <https://www.technologyreview.es/s/10924/como-se-produce-el-sesgo-algoritmico-y-por-que-es-tan-dificil-detenerlo>.

—. **2019b.** Facebook's ad-serving algorithm discriminates by gender and race. *MIT Technology Review*. [En línea] 5 / Abril / 2019b. [Data: 4 / Febrer / 2021.] <https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/>.

HOLLEY, Peter. 2018. Elon Musk's nightmarish warning: AI could become 'an immortal dictator from which we would never escape'. *The Washington Post*. [En línea] 6 / Abril / 2018. [Data: 11 / Març / 2021.] <https://www.washingtonpost.com/news/innovations/wp/2018/04/06/elon-musks-nightmarish-warning-ai-could-become-an-immortal-dictator-from-which-we-would-never-escape/>.

HUMAN RIGHTS WATCH. 2018. *Atender la llamada. Un imperativo moral y legal prohibir los robots asesinos*. 2018.

—. **2016.** *Killer robots and the Concept of Meaningful Human Control*. 2016.

—. **2016.** *Making the Case. The Dangers of Killer Robots and the Need for a Preemptive Ban*. 2016.

—. **2020.** *Stopping Killer Robots. Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control*. 2020.

JORGE URBINA, Julio. 2000. *Protección de las víctimas de los conflictos armados, naciones unidas y derecho internacional humanitario*. València : Tirant lo Blanch, 2000.

KANTARCI, Atakan. 2021. Bias in AI: What it is, Types & Examples, How & Tools to fix it. *AIMultiple*. [En línia] 2 / Febrer / 2021. [Data: 4 / Febrer / 2021.] <https://research.aimultiple.com/ai-bias/>.

KERR, Ian i SZILAGYI, Katie. 2016. Asleep at the switch? How killer robots become a force multiplier of military necessity. [autor llibre] Ryan CALO, A. Michael FROOMKIN i Ian KERR. *Robot Law*. Cheltenham : Edward Elgar Publishing, 2016.

KOTTASOVA, Ivana. 2017. Europe calls for mandatory 'kill switches' on robots. *CNN*. [En línia] 12 / Gener / 2017. [Data: 27 / Abril / 2021.] <https://money.cnn.com/2017/01/12/technology/robot-law-killer-switch-taxes/index.html>.

LEE, Peter. 2018. Drones will soon decide who to kill. *The conversation. Rigor académico, oficio periodístico*. [En línia] 11 / Abril / 2018. [Data: 18 / Febrer / 2021.] <https://theconversation.com/drones-will-soon-decide-who-to-kill-94548>.

MARTÍNEZ QUIRANTE, Roser i RÓDRIGUEZ ÁLVAREZ, J. 2018. *Inteligencia artificial y armas letales autónomas*. Gijón : Ediciones TREA, 2018.

MARTÍNEZ QUIRANTE, Roser i RÓDRIGUEZ ÁLVAREZ, Joaquín. 2019. *Towards a new AI race. The challenge of lethal autonomous weapons systems (LAWS) for the United Nations*. Pamplona : Editorial Aranzadi, 2019.

MARTÍNEZ, Naroa i MATUTE, Helena. 2020. DISCRIMINACIÓN RACIAL EN LA INTELIGENCIA ARTIFICIAL. *Ethic*. [En línia] 10 / Setembre / 2020. [Data: 3 / Febrer / 2021.] <https://ethic.es/2020/09/discriminacion-racial-en-la-inteligencia-artificial-sesgos/>.

NARRILLOS, Estefanía. 2020. El Parlamento muestra el camino para la normativa sobre inteligencia artificial. *Noticias Parlamento Europeo*. [En línia] 21

/ Octubre / 2020. [Data: 4 / Febrer / 2021.]
<https://www.europarl.europa.eu/news/es/press-room/20201016IPR89544/el-parlamento-muestra-el-camino-para-la-normativa-sobre-inteligencia-artificial>.

NAVAS NAVARRO, Susana, et al. 2017. *Inteligencia artificial. Tecnología. Derecho*. Valencia : Tirant lo Blanch, 2017.

O'CARROLL, Brodie. 2017. What are the three types of AI? A guide to narrow, general, and super artificial intelligence. *Codebots*. [En línea] 24 / Octubre / 2017. [Data: 3 / Febrer / 2021.] <https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible>.

PARLAMENT EUROPEU I CONSELL (art.4.4). 2016. *REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que deroga[...]*. Bruselas : s.n., 2016. DOUE-L-2016-80807.

PARLAMENT EUROPEU. 2021. Regulación de la inteligencia artificial en la UE: la propuesta del Parlamento. *NOTICIAS PARLAMENTO EUROPEO*. [En línea] 21 / Gener / 2021. [Data: 4 / Febrer / 2021.]

—. **2017.** *Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica*. Estrasburg : s.n., 2017. 2015/2103(INL).

PASQUALE, Frank. 2020. ‘Machines set loose to slaughter’: the dangerous rise of military AI. *The Guardian*. [En línea] 15 / Octubre / 2020. [Data: 27 / Abril / 2021.] <https://www.theguardian.com/news/2020/oct/15/dangerous-rise-of-military-ai-drone-swarm-autonomous-weapons>.

PC PRO. 2015. Killer robots: who's responsible? *PC Pro; London*. [En línea] Octubre / 2015. [Data: 28 / Abril / 2021.] <https://search-proquest-com.are.uab.cat/magazines/killer-robots-whos-responsible/docview/1711126926/se-2?accountid=15292>.

PELEGRÍ, Jordi. 2020. INTELIGENCIA ARTIFICIAL Y ROBÓTICA: LA BÚSQUEDA DE LA PERFECCIÓN EN LA PRODUCCIÓN. *Universal Robots*. [En línia] 11 / Març / 2020. [Data: 14 / Febrer / 2021.] <https://blog.universal-robots.com/es/inteligencia-artificial-y-robotica>.

PINO DíEZ, Raúl, GÓMEZ GÓMEZ, Alberto i ABAJO MARTÍNEZ, Nicolás de. 2001. *Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva*. Universidad de Oviedo : Servicio de publicaciones, 2001.

RIPIPSA. 2020. ROBOTS AUTÓNOMOS: QUÉ SON, CÓMO FUNCIONAN Y QUE VENTAJAS OFRECEN. *RIPIPSA, S.A.* [En línia] 6 / Agost / 2020. [Data: 14 / Febrer / 2021.] <https://ripipsacobots.com/robots-autonomos/>.

RODRÍGUEZ-VILLASANTE Y PRIETO, José Luis. 2007. *ARMAS NO LETALES INHABILITANTES Y DERECHO INTERNACIONAL HUMANITARIO*. Bogotá : Cruz Roja, 2007.

RUIZ, Enrique i ARAGÓN SÁNCHEZ, Teresa. 2017. Diferencias entre responsabilidad civil subjetiva y objetiva. *LegalToday. Por y para profesionales del Derecho*. [En línia] 12 / Juliol / 2017. [Data: 28 / Abril / 2021.] <https://www.legaltoday.com/practica-juridica/derecho-civil/civil/diferencias-entre-responsabilidad-civil-subjetiva-y-objetiva-2017-07-12/>.

SOCIETY OF AUTOMOTIVE ENGINEERS. 2018. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. 2018. J3016_201806.

SPARROW, Robert. 2017. Killer Robots. *Journal of Applied Philosophy*. 1 / Novembre / 2017, Vol. 24.

TAYLOR, Isaac. 2020. Who Is Responsible for Killer Robots? Autonomous Weapons, Group Agency, and the Military-Industrial Complex. *Journal of Applied Philosophy*. 3 / Novembre / 2020, p. 15.

TECH2 NEWS STAFF. 2017. FACEBOOK RESEARCHERS SHUT DOWN AI BOTS THAT STARTED SPEAKING IN A LANGUAGE UNINTELLIGIBLE TO HUMANS. *TECH2*. [En línia] Firstpost, 1 / Agost / 2017. [Data: 4 / Febrer / 2021.] <https://www.firstpost.com/tech/news-analysis/facebook-researchers-shut-down-ai-bots-that-started-speaking-in-a-language-unintelligible-to-humans-3876197.html>.

TECHNICAL COMMITTEE FOR COGNITIVE ROBOTS. IEEE Robotics & Automation Society. [En línia] [Data: 3 / Gener / 2021.] <https://www.ieee-ras.org/cognitive-robotics>.

WALCH, Kathleen. 2020. Will There Be A 'Kill Switch' For AI? *Forbes*. [En línia] 5 / Març / 2020. [Data: 27 / Abril / 2021.] <https://www.forbes.com/sites/cognitiveworld/2020/03/05/will-there-be-a-kill-switch-for-ai/?sh=68aa9d952ef5>.

WALSH, Toby. 2020. Killer robots. La intel·ligència artificial i les armes autònomes. 19 / Febrer / 2020, NÚM. 48 INTEL·LIGÈNCIA ARTIFICIAL.

YAKIMOVA, Yasmina. 2021. Guidelines for military and non-military use of Artificial Intelligence. *News European Parliament*. [En línia] 20 / Gener / 2021. [Data: 4 / Febrer / 2021.] <https://www.europarl.europa.eu/news/en/press-room/20210114IPR95627/guidelines-for-military-and-non-military-use-of-artificial-intelligence>.

